

1 1

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
1. R57006	*** 3 standard deviations	above mean	***		
2. R32437	N-terminal fragment of human	10	4	4	3.70 0
3. R31647	Human thrombospondin type I r	10	4	6	3.70 0
4. P80008	Sequence deduced from human p	11	4	4	3.70 0
5. R03394	Synthetic peptide Sp-23 with	11	4	4	3.70 0
6. R05524	Fragment of pertussis toxin S	11	4	4	3.70 0
7. P10340	Tumour necrosis factor derive	12	4	4	3.70 0
8. R06485	Somatostatin deriv. A50.	13	4	4	3.70 0
	N-terminal of plasma protein	13	4	4	3.70 0

425

9. R41498 TNF inhibitory peptide X.
10. R57135 Recombinant mature xenoxin-1
11. R51437 Circumsporozoite protein regi
12. R09375 Sequence of peptide TPI used
13. R91752 Antigenic N-terminal peptide
14. R32709 N terminal of haematopoietic
15. R47008 IFN-alpha receptor position 2
16. R97787 Sequence encoded by a genetic
17. R51436 Circumsporozoite protein regi
18. R30935 Peptide 43 corresponding to 1
19. R26535 Sequence of peptide P3 which
20. R26534 Sequence of peptide P2 which
21. R13689 Scatter factor N-terminal.
22. R13688 Papilloma virus type 16 L1 pe
23. R04980 Papilloma virus type 16 L1 pe
24. R04981 Thrombospondin peptide p11.
25. R13629 HIV epitope #58.
26. R47725 Hib OMP P2 peptide OMP2-7 (17
27. R40077 Hepatocyte growth factor beta
28. R10144 Halystatin fragment.
29. R42872 Sequence of peptide TPQ used
30. R09376 Arginine deiminase.
31. R03358 Exon XIII of human hepatocyte
32. R25690 Amphiphilic peptide #119 used
33. R35388 Cecropin B.
34. P94254 Spider venom calcium channel
35. R53573 Amino acid sequence of Mod-S
36. P91331 Amino acid sequence of SB-37
37. P91330 Amino acid sequence of cecrop
38. P91063 Semisynthetic Cecropin B gene
39. P94255 Hyalophora cecropia Cecropin
40. R53353 Lytic peptide with proliferat
41. R07734 pAGbeta2-101 platelet aggrega
42. R46945 pAGbeta107 platelet aggrega
43. R46944 pAGbeta106 platelet aggregati
44. R46943
45. R46942

1. US-08-121-713B-1 (1-6) N-terminal fragment of human histo-blood group A t
R57006 ID R57006 standard; peptide; 10 AA.
AC R57006;
DT 15-FEB-1995 (first entry)
DE N-terminal fragment of human histo-blood group A transferase.
KW Blood; group; determinant; antigen; erythrocyte; oligosaccharide;
KW glycoconjugate; glycosphingolipid; glycoprotein; glycosyltransferase;
KW transferase.
OS Homo sapiens.
PN US5326857-A.
PD 05-JUL-1994.
PF 31-AUG-1989; 402695.
PR 31-AUG-1989; US-402695.
PR 29-AUG-1991; US-752101.
PA (BIOM-) BIOMEMBRANE INST.
PI Clausen H, Hakomori S, White T, Yamamoto F;
DR WPI; 94-217098/26.
PT Isolated DNA molecules - encode human histo-blood groups A-, B-
and O-glycotransferases

PS Example 2; Column 41-42; 63pp; English.
CC The histo-blood group ABH determinants are major allogeneic antigens
CC in both erythrocytes and tissues of humans. They generally
CC constitute peripheral parts of the oligosaccharide chains of
CC glycoconjugates i.e. linked to lipids (glycosphingolipids) or to
CC proteins (glycoproteins). It was proposed that the A and B
CC phenotypes were associated with glycosyltransferases that converted
CC the H substance associated with the O phenotype to A and B
CC respectively, through the addition of alphanal-3-N-acetylgalactosamine
CC or alphanal-3-galactosyl residues to the H antigen Fuc-alpha-2Gal-
CC betaal-R. Hence, the primary products of the histo-blood group A
CC and B genes are the respective glycosyltransferases. This is a
CC fragment of the A group transferase. See also R56995-R57010.
SQ Sequence 10 AA;
SQ 2 A; 0 R; 1 N; 1 D; 0 B; 0 C; 1 Q; 1 E; 0 Z; 1 G; 0 H;
SQ 1 I; 0 L; 0 K; 0 M; 0 F; 0 P; 0 S; 0 T; 0 W; 0 Y; 2 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.70
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
XCXNXI
I I
VDQANGIEAV
X X 10
2. US-08-121-713B-1 (1-6) Human thrombospondin type I repeat-derived peptide
R32437 ID R32437 standard; peptide; 10 AA.
AC R32437;
DT 10-JUN-1993 (first entry)
DE Human thrombospondin type I repeat-derived peptide #8.
KW Thrombosis; coagulation; heparin binding inhibitor; type I repeat.
OS Synthetic.
PN US7801812-A.
PD 15-DEC-1992.
PF 06-DEC-1991; 801812.
PR (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Roberts DD;
DR WPI; 93-067439/08.
PT New sulphated glyco-conjugate binding peptide(s) - from type 1
PT repeats of human thrombospondin, preventing interaction of the
PT glyco-conjugates with adhesion molecules, growth factors, etc.
PS Disclosure; Page 13; 64pp; English.
CC This peptide was obtained from the adhesive glycoprotein thrombo-
CC spondin. It is adjacent to an active, preferred peptide of the
CC invention which lacks the cluster of basic amino acids which fit
CC the consensus sequences of many heparin-binding proteins but which
CC have binding constants of approximately 10 to 100-fold higher than
CC the proteins having the basic consensus sequences. The pref.
CC peptides all contain a subsequence WSXW (X=P, E, H, A, S) with a
CC substantial lack of an electrical charge. They act as inhibitors of
CC heparin- or related sulphated glycoconjugate-binding to adhesion
CC molecules, growth factors, etc. This flanking sequence does not
CC have inhibitory activity.
SQ Sequence 10 AA;

SQ 0 A; 1 R; 1 N; 0 D; 0 B; 1 C; 2 Q; 0 E; 0 Z; 2 G; 0 H;
 SQ 1 I; 0 L; 0 K; 0 M; 0 F; 0 P; 1 S; 1 T; 0 W; 0 Y; 0 V;
 Initial Score = 4 Optimized Score = 6 Significance = 3.70
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 XCXNXI
 | |
 TSCNGIQQR
 X X 10

3. US-08-121-713B-1 (1-6) Sequence deduced from human papillomavirus (HPV) o

ID P91647 standard; peptide; 11 AA.
 AC P91647;
 DT 11-MAR-1992 (first entry)
 DE Sequence deduced from human papillomavirus (HPV) open reading
 DE frame.
 KW Diagnosis; wart; condylomas; cervical cancer; antigen.
 OS Human papillomavirus.
 PN EP-344940-A.
 PD 06-DEC-1989.
 PF 15-MAY-1989; 304874.
 PR 16-MAY-1988; US-194407.
 PR 13-MAR-1989; US-323614.
 PA (SRI-) SCRIPPS CLINIC RES.
 PI Dillner J, Lerner RA, Smith R, Parks DE;
 DR WPI; 89-358324/49.
 PT Antigenic polypeptide(s) useful as immunogens - for producing
 PT antibodies to papillomavirus agent proteins
 PS Claim 3; Page 49; 60pp; English.
 CC The peptides of the invention are useful for generating antibodies
 CC that immunoreact with papillomavirus latent protein. Inocula
 CC typically contain polypeptide or protein concns. of 10mcg to 500mcg
 CC per dose (pref. 50mcg-50mg).
 SQ Sequence 11 AA;
 SQ 0 A; 0 R; 1 N; 0 D; 0 B; 0 C; 0 Q; 1 E; 0 Z; 1 G; 0 H;
 SQ 1 I; 0 L; 0 K; 0 M; 1 F; 0 P; 0 S; 3 T; 0 W; 0 Y; 3 V;
 Initial Score = 4 Optimized Score = 4 Significance = 3.70
 Residue Identity = 66% Matches = 2 Mismatches = 1
 Gaps = 0 Conservative Substitutions = 0

X X
 XCXNXI
 | |
 NGIVTVFVTE
 X X 10

4. US-08-121-713B-1 (1-6) P80008 Synthetic peptide Sp-23 with residues 2-11 deduced

ID P80008 standard; protein; 11 AA.
 AC P80008;
 DT 14-SEP-1990 (first entry)

DE Synthetic peptide Sp-23 with residues 2-11 deduced from the gag sequence
 DE of a human proviral locus endogenous retrovirus-1
 KW Synthetic peptide Sp-23; endogenous retrovirus-1; tumour detection;
 OS chromosome 18.
 PN erv-1 provirus/Homo sapiens.
 PD US4777127-A.
 PF 11-OCT-1988.
 PR 30-SEP-1985; 781478.
 PR 30-SEP-1985; US-781478.
 PA (LABS-) Labavstem Oy.
 PI Jukka S, Antti V;
 DR WPI; 88-307164/43.
 PT Purified human retro-virus-related peptide prods. -
 PT useful as specific reagents for detection and treatment of
 PT tumours, such as renal cell adenocarcinoma and choriocarcinoma
 PS ; Table 1, pages 3 and 4; 6pp; English.
 CC A defective, endogenous provirus was isolated from a human recombinant
 CC DNA library by using an endogenous chimpanzee retroviral pol fragment as
 CC probe. The structure of this fragment is highly related to that of BaEV
 CC This genome, termed HC-20 (or endogenous retrovirus-1; erv-1), has been
 CC assigned to human chromosome 18. A decapeptide, deduced from the gag
 CC portion of erv-1 has been isolated. From this, an undecapeptide (Sp-23)
 CC which contains the decapeptide as residues 2 through to 11 has been
 CC synthesized. A polyclonal antibody raised against the undecapeptide in
 CC rabbits has detected a human retrovirus-related Mr75,00 protein which
 CC contains the decapeptide. Polyclonal antibodies have been raised against
 CC the decapeptide, Sp-23 and the Mr 75,000 protein. These antibodies are
 CC used as specific reagents for the detection of tumours such as renal cel
 CC adenocarcinoma and choriocarcinoma, among others, and the placental
 CC disorders including blighted ova, hydatidiform and destructive moles. The
 CC sequence of synthetic peptide (sp-23) fits 6 of 10 with the BaEV p30 and
 CC 7 of 10 with the Mo-MuLV p30 sequence.

SQ Sequence 11 AA;
 SQ 0 A; 1 R; 1 N; 0 D; 0 B; 1 C; 1 Q; 2 E; 0 Z; 0 G; 0 H;
 SQ 0 I; 1 L; 0 K; 0 M; 1 F; 1 P; 1 S; 0 T; 0 W; 1 Y; 0 V;
 Initial Score = 4 Optimized Score = 4 Significance = 3.70
 Residue Identity = 40% Matches = 2 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 XCXNXI
 | |
 CENPSQFYERL
 X X 10

5. US-08-121-713B-1 (1-6) R03394 Fragment of pertussis toxin S1 subunit.

ID R03394 standard; protein; 11 AA.
 AC R03394;
 DT 31-JUL-1990 (first entry)
 DE Fragment of pertussis toxin S1 subunit.
 KW Whooping cough; subunit S1; vaccine.
 OS Bordetella pertussis.
 PN W09001494-A.
 PD 22-FEB-1990.
 PF 31-JUL-1989; 03298.
 PR 2-AUG-1988; US-227372.

PA (STRD) Leland Stanford Jr Univ.
PI Steinman L, Oksenberg JR, Schoolnik GK, Judd AK;
DR WPI; 90-083479/11.
PT Polypeptide(s) useful as Bordetella pertussis vaccines -
PT comprise specified sequence(s) of SI subunit.
PS Disclosure; pp; English.
CC Useful as a vaccine for whooping cough, free from the side effects
CC associated with the intact SI subunit.
SQ Sequence 11 AA;
SQ 0 A; 1 R; 1 N; 0 D; 0 B; 0 C; 0 Q; 1 E; 0 Z; 2 G; 1 H;
SQ 1 I; 0 L; 0 K; 0 M; 0 F; 0 P; 0 S; 2 T; 0 W; 1 Y; 1 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.70
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| |
RVYHNGITGET
X X 10

6. US-08-121-713B-1 (1-6)
R05524 Tumour necrosis factor derived peptide.

ID R05524 standard; protein; 12 AA.
AC R05524;
DT 24-OCT-1990 (first entry)
DE Tumour necrosis factor derived peptide.
KW Tumour necrosis factor; TNF; neoplastic disease; autoimmune
KW disease; infection; inflammation; transplant rejection.
OS Synthetic.

FH Key Location/Qualifiers
FT Modified-site 4..4
FT /label=Hcy
FT Disulfide-bond 4..10
DE3841768-A.
PN DE3841768-A.
PD 13-JUN-1990.
PF 12-DEC-1988; 841768.
PR 12-DEC-1988; DE-841768.
PA (BADI) BASF AG.
PI Bohm HJ, Daum L, Haupt A, Schmied B, Walker N, Zechel JC;
DR WPI; 90-186584/25.
PT New tumour necrosis factor derived peptide(s) - for treating or
PT preventing neoplastic and auto-immune diseases, infection, inflammation
PT and transplant rejection reaction.
PS Example 60; Page 10; 17pp; German.
CC To residue N1 is attached Ac and to residue D12 NH2.
CC This peptide is an example of a highly generic sequence of the
CC formula X-A-Gly-B-Y
CC A= Asp, His or Asn;
CC B= Val, Met or Phe;
CC X= G-NH-CHM-CO, G-NH-CHM-CO-W, G-R-NH-CHM-CO-W;
CC Y= Z, NH-CHQ-COZ, V-NH-CHQ-COZ, NH-CHQ-CO-U-Z or V-NH-CHQ-CO-U-Z;
CC G= H or an amino protecting group;
CC Z= OH, NH2 or carboxy protecting group; or G and Z together are a
CC covalent bond or the gp. CO(CH2)anH; a=1-12;
CC R, U, V and W= peptide chains of 1-4 naturally occurring alpha aminoacids;
CC M and Q= H, isopropyl, sec-butyl, phenyl, 1-hydroxy-ethyl, 3-indolyl- or

CC 4-imidazolyl-methyl or (CH2)bT; b=1-6;
CC T= OH, MeO, MeS, isopropyl, phenyl (opt. 4-OH, substd), mercapto,
CC amino, carboxy, carbamoyl or guanidino; or
CC M ans Q together are (CH2)c-S-S-(CH2)d, (CH2)eCO NH-(CH2)f or
CC (CH)2eNH CO(CH2)gNH CO(CH2) f; c and d=1-4; e and f=1-6; g=1-12.
CC The peptide is a low mol. wt. deriv. of TNF.
CC See also DE3841753-55, DE3841759, DE3841761-64, DE3841767-68.
SQ Sequence 12 AA;
SQ 2 A; 2 N; 1 D; 0 B; 2 C; 0 Q; 1 E; 0 Z; 1 G; 0 H;
SQ 0 I; 1 L; 0 K; 0 M; 0 F; 0 P; 0 S; 0 T; 0 W; 0 Y; 1 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.70
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| |
NALCANGVECRD
X X 10

7. US-08-121-713B-1 (1-6)
P10340 Somatostatin deriv. A50.

ID P10340 standard; peptide; 13 AA.
AC P10340;
DT 15-DEC-1992 (first entry)
DE Somatostatin deriv. A50.
KW Growth Hormone secretion; GH; diabetes mellitus; angiopathy;
KW acromegaly; diagnosis.
OS Synthetic.

FH Key Location/Qualifiers
FT Modified-site 1
FT /note=*(4Cl)Phe*
FT Modified-site 3
FT /label=Nle
FT Misc difference 14
FT /note=*(Cys-NH-(CH2)2-OH, Cys-NH-(CH2)3-OH or
FT as amide formed with 3-amino-butyrolactone*
FN CH-621770-A.
PD 27-FEB-1981.
PF 11-SEP-1980; 125375.
PR 23-FEB-1976; CH-002175.
PA (SANO) SANDOZ AG.
PI Sandrin E, Bauer W;
DR WPI; 81-21515D/13 (21515D).
PT Somatostatin derivs. prodn. - useful for treating diabetes,
PT acromegalia and angiopathia
PS Example 1; Page 6-7; 8pp; German.
CC This peptide is an example of a generic formula for somatostatin
CC derivs. which inhibit secretion of growth hormone and are useful to
CC treat diabetes mellitus, acromegaly, angiopathy and in diagnosis.
CC See P10308-P10348.
SQ Sequence 13 AA;
SQ 0 A; 0 R; 1 N; 0 D; 0 B; 2 C; 0 Q; 0 E; 0 Z; 0 G; 0 H;
SQ 0 I; 0 L; 1 K; 0 M; 4 F; 0 P; 1 S; 2 T; 1 W; 0 Y; 0 V;
SQ 1 Others;
Initial Score = 4 Optimized Score = 4 Significance = 3.70

Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
|||
FCXNFWKTFSTC
X X 10

8. US-08-121-713B-1 (1-6)
R06485 N-terminal of plasma protein which binds activated

ID R06485 standard; peptide; 13 AA.
AC R06485;
DT 03-JAN-1991 (first entry)
DE N-terminal of plasma protein which binds activated C4.
KW Complement; 120 kDa protein.
OS Homo sapiens.
PN W09008770-A.
PD 09-AUG-1990.
PF 02-FEB-1990; U00716.
PR 02-FEB-1989; US-305458.
PA (USDC) US SEC OF COMMERCE.
PI Hammer CH, Jacobs RM, Frank MM;
DR WPI; 90-260889/34.
PT New purified single-chain plasma protein - binds to activated C4
PT and C3 of the classical complement pathway.
PS Claim 2; Page 28; 40pp; English.
CC The new protein, of which this sequence is the N-terminal, is a
CC single chain plasma protein of 120 kD which inhibits lytic
CC functional activity of the classical complement pathway.
CC Quantitative Manchini analysis has identified about 300 ul of the
CC protein in plasma and serum. The molecule contains approx. equal
CC amts. of N- and O-linked CHO (14%). It is cleaved by Kallikrein
CC into 85 and 35 kD fragments; the 85 kD fragment is subsequently
CC cleaved into fragments of 60 and 25 kD. Studies show that a mixt.
CC of the fragments, but not the intact protein, has a potent vaso-
CC dilating activity in guinea pigs. The protein is believed to be
CC a new complement regulatory factor and shares several character-
CC istics with the C2 protein.
SQ Sequence 13 AA;
SQ 0 A; 0 R; 1 N; 2 D; 0 B; 0 C; 0 Q; 1 E; 0 Z; 1 G; 0 H;
SQ 2 I; 1 L; 1 K; 0 M; 0 F; 0 P; 1 S; 1 T; 0 W; 1 Y; 1 V;

Initial Score = 4 Optimized Score = 4 Significance = 3.70
Residue Identity = 40% Matches = 2 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
|||
EKNGIDYSLTVD
X X 10

9. US-08-121-713B-1 (1-6)
R41498 TNF inhibitory peptide X.

ID R41498 standard; peptide; 14 AA.

AC R41498;
DT 23-FEB-1994 (first entry)
DE TNF inhibitory peptide X.
KW Tumour necrosis factor; TNF; inhibition; solid phase synthesis; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT Disulfide bond 3..12
FT /note= "Optional di-sulphide bond"

PN J05194594-A.
PD 03-AUG-1993.
PF 21-JAN-1992; 029044.
PR 21-JAN-1992; JP-029044.
PA (SAGA) SAGAMI CHEM RES CENTRE.
DR WPI; 93-282916/36.
PT TNF inhibitory novel peptide(s) - include N-terminal amino Gp.
PT which is opt. modified with acetyl, t-butoxy-carbonyl or
PT benzyl-oxy-carbonyl Gp. and C-terminal carboxy Gp. is opt.
PT amidated
PS Claim 1; Page 6; 8pp; Japanese.
CC The sequences given in R41489-99 are tumour necrosis factor (TNF)
CC inhibitory peptides. They may optionally be modified at the N-
CC terminal with an acetyl, t-butoxycarbonyl or benzylloxycarbonyl, and
CC at the C-terminal they are optionally amidated. These peptides are
CC produced by solid phase synthesis methods and may be produced at low
CC cost.

SQ Sequence 14 AA;
SQ 0 A; 0 R; 1 N; 0 D; 0 B; 2 C; 1 Q; 1 E; 0 Z; 1 G; 1 H;
SQ 0 I; 3 L; 0 K; 0 M; 0 F; 0 P; 2 S; 1 T; 0 W; 0 Y; 1 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.70
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
|||
SLCLNGTVHLSCQE
X X 10

10. US-08-121-713B-1 (1-6)
R57135 Recombinant mature xenoxin-1 N-terminal sequence.

ID R57135 standard; Protein; 15 AA.
AC R57135;
DT 22-FEB-1995 (first entry)
DE Recombinant mature xenoxin-1 N-terminal sequence.
KW Defensin A; pro-sequence; heterologous protein production;
KW fleshfly; yeast expression vector; yeast alpha-mating factor;
KW pre-BGL2; xenoxin-1; Xenopus laevis.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc difference 3
FT /note= "not identified by sequencing method"
FT Misc difference 14
FT /note= "not identified by sequencing method"
PN EP-607080-A.
PD 20-JUL-1994.
PF 11-JAN-1994; 400062.
PR 11-JAN-1993; FR-000171.

PA (TRGE) TRANSGENE SA.
PI Achstetter T;
DR WPI; 94-226998/28.
PT Cassette for expression and synthesis of mature heterologous
PT protein in yeast - included pre sequence and pro sequence from an
PT insect defensin precursor, also new expression vectors and
PT transformed hosts
PS Example 6; Page 30; 39pp; French.
CC Expression vectors were constructed for recombinant expression of
CC xenoxin-1 in transformed yeast hosts. The vectors comprised the
CC sequence of pre BGL2, the pro sequence yeast alaph-mating factor
CC and the sequence coding for mature xenoxin-1, under the
CC control of a modified alaph-WF promoter. The construction
CC procedure for preparing the vectors included PCR amplification of
CC a sequence coding for mature xenoxin-1 (having the amino acid
CC sequence R57134) using the primers Q69212 and Q69213. The amplified
CC product had an EagI site at the 3'-end and a SalI site at the 3'-end.
CC The N-terminus of the recombinantly produced protein was sequenced
CC and was found to correspond to the expected N-terminal sequence for
CC xenoxin-1 (see R57135).
SQ Sequence 15 AA;
SQ 1 A; 0 R; 2 N; 0 D; 0 B; 0 C; 2 Q; 0 E; 0 Z; 1 G; 0 H;
SQ 1 I; 2 L; 2 K; 1 M; 0 F; 0 P; 0 S; 0 T; 0 W; 0 Y; 1 V;
SQ 2 Others;

Initial Score = 4 Optimized Score = 4 Significance = 3.70
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| |
LXVNLQANGIRMXQ
X 10

11. US-08-121-713B-1 (1-6) Circumsporozoite protein region II+ peptide #11.
R51437

ID R51437 standard; peptide; 15 AA.
AC R51437;
DT 26-OCT-1994 (first entry)
DE Circumsporozoite protein region II+ peptide #11.
KW Region II+; circumsporozoite; CS; Plasmodium; inhibitor; binding;
KW receptor; hepatocytes; malaria-susceptible; mammal; infection;
KW antibody; liver; human hepatocyte cell line; HepG2.
OS Plasmodium sp.
PN W09406464-A.
PD 31-MAR-1994.
PF 17-SEP-1993; U08800.
PR 17-SEP-1992; US-947033.
PA (UNY) UNIV NEW YORK STATE.
PI Cerami C, Frevet U, Nussenzweig V, Sinns P;
DR WPI; 94-118161/14.
PT Peptide(s) corresp. to Region II+ of circumsporozoite - are used
PT for inhibiting hepatocyte invasion by malarial sporozoites for
PT preventing malaria infection
PS Claim 10; Page 33; 42pp; English.
CC The sequences given in R51427-35 are peptide fragments derived from
CC region II+ of the circumsporozoite (CS) protein derived from various

CC Plasmodium species. These peptides fragments may be used as an
CC inhibitor for the binding of CS polypeptides to receptors of
CC hepatocytes from malaria-susceptible mammals. These peptides can be
CC administered to malaria susceptible mammals to prevent infection,
CC and they can be used to produce antibodies which can be used to
CC prevent infection. The peptides were tested for inhibition of CS
CC binding to liver sections or to the human hepatocyte cell line HepG2.
SQ Sequence 15 AA;
SQ 0 A; 1 R; 1 N; 0 D; 0 B; 2 C; 1 Q; 0 E; 0 Z; 2 G; 0 H;
SQ 2 I; 0 L; 1 K; 0 M; 0 F; 1 P; 1 S; 1 T; 0 W; 0 Y; 2 V;
Initial Score = 4 Optimized Score = 6 Significance = 3.70
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| |
PCSVTCGNGIQVIRK
X 10

12. US-08-121-713B-1 (1-6) Sequence of peptide TPI used to isolate onco-devel
R09375

ID R09375 standard; Protein; 15 AA.
AC R09375;
DT 26-MAR-1992 (first entry)
DE Sequence of peptide TPI used to isolate onco-developmentally
DE regulated alpha-N-acetyl-galactosaminyltransferase (AGT).
KW Tumour diagnosis; tumour cell; enzyme; fibronectin;
PN EP-354652-A.
PD 14-FEB-1990.
PF 22-JUN-1989; 306283.
PR 12-AUG-1988; US-231576.
PA (BIOM-) BIOMEMBRANE INST.
PA (HUTC-) HUTCHINSON F CANCER RES.
PI Hakomori SI;
DR WPI; 90-046346/07.
PT Onco-developmentally regulated n-acetyl
PT galactosaminyl-transferase - isolated from onco-foetal tissue
PT membrane fraction and used in detection of tumour cells
PS Claim 3; Page 13; 16pp; English.
CC The inventors claim a method for the isolation of AGT as a component
CC of a particulate membrane fraction sepd. from cell and tissue
CC homogenates. The prefd. method is by contacting onco foetal tissue
CC with an immobilised polypeptide substrate comprising
CC VTY (R09374); TPI (R09375); TPQ (R09376); or a normal fibronectin to
CC form an affinity complex and then eluting the enzyme.
SQ Sequence 15 AA;
SQ 0 A; 0 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 3 G; 1 H;
SQ 1 I; 0 L; 0 K; 0 M; 1 F; 2 P; 0 S; 3 T; 0 W; 1 Y; 1 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.70
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| |

TPFVTHPGYDTGNGI
10 X

13. US-08-121-713B-1 (1-6)
P91752 Antigenic N-terminal peptide of human fibrinogen g

ID P91752 standard; protein; 15 AA.
AC P91752;
DT 17-MAY-1990 (first entry)
DE Antigenic N-terminal peptide of human fibrinogen gamma chain.
KW N-terminal; antigenic peptide; fibrinogen; human leukocyte elastase;
OS Homo sapiens.
PN EP-345906-A.
PD 13-DEC-1989.
PF 07-JUN-1989; 201460.
PR 10-JUN-1988; US-205416.
PR 10-JUN-1988; US-205417.
PR 10-JUN-1988; US-205418.
PI (MERI) Merck and Co. Inc.
PI Dahlgren ME, Mumford RA, Boger JS, Davies DTP;
DR WPI; 89-365672/50.
PT New elastase-induced fibrinogen cleavage peptide(s) - for prodn. of
PS Disclosure; page 8; 33pp; English.
CC connective tissue disease; gamma chain;
CC new diagnostic antibodies.
CC The peptide represents residues 358-372 of human fibrinogen (hF) gamma
CC chain. Tyrosines at positions 12 and 15 below are a tentative
CC assignment. The human leukocyte elastase (HLE) cleavage prod.
CC (AAs 1-4 below) is claimed (no. 4) and is used to raise monoclonal
CC antibodies. These are useful for rapid detection of HLE-induced cleavage
CC of hF, eg in diagnosis of connective tissue destructive diseases such
CC as emphysema, chronic bronchitis, cystic fibrosis, arthritis, psoriasis,
CC or atherosclerosis.
CC See also P91744-P91751.
SQ Sequence 15 AA;
SQ 1 A; 0 R; 2 N; 1 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 2 G; 0 H;
SQ 2 I; 0 L; 0 K; 0 M; 0 F; 1 P; 1 S; 2 T; 0 W; 3 Y; 0 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.70
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
|
STPGYDNGIIVATY
X 10

14. US-08-121-713B-1 (1-6)
R32709 N terminal of haematopoietic stem cell multiplier.

ID R32709 standard; Protein; 16 AA.
AC R32709;
DT 16-JUN-1993 (first entry)
DE N terminal of haematopoietic stem cell multiplier.
KW Bone marrow deficiencies; cancer therapy; tumour; carcinoma;
PN W09303061-A.

PD 18-FEB-1993.
PF 24-JUL-1992; J00949.
PR 26-JUL-1991; JP-187470.
PR 26-JUL-1991; JP-187481.
PA (TORA) TORAY IND INC.
PI Kawano G, Kojima K, Koniya A, Kubo T, Nakahata T;
PI Sano E, Sudot, Tanaka R;
DR WPI; 93-076441/09.
PT Haematopoietic stem cell multiplier comprising IL-3 and IL-7 -
PT used for treatment and prevention of bone marrow disorders e.g.
PT after cancer therapy or bone marrow transplants
PS Disclosure; Page 36; 90pp; Japanese.
CC This sequence is the N terminal sequence of a haematopoietic stem
CC cell multiplier. *X* = any amino acid.
SQ Sequence 16 AA;
SQ 0 A; 0 R; 2 N; 0 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 2 G; 0 H;
SQ 2 I; 0 L; 1 K; 1 M; 0 F; 1 P; 0 S; 2 T; 0 W; 0 Y; 3 V;
SQ 2 Others;

Initial Score = 4 Optimized Score = 4 Significance = 3.70
Residue Identity = 40% Matches = 2 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
|
VVGIFTXTNIGXW
X X 10

15. US-08-121-713B-1 (1-6)
R47008 IFN-alpha receptor position 271-287.

ID R47008 standard; Protein; 17 AA.
AC R47008;
DT 16-SEP-1994 (first entry)
DE IFN-alpha receptor position 271-287.
KW Naturally-occurring; immunomodulatory protein; human; therapy; class I;
KW major histocompatibility complex; class II; allotype; type I diabetes;
KW autoimmune disease; rheumatoid arthritis; T-cell-mediated response;
KW multiple sclerosis; transplant rejection; vaccine; MHC.
OS Homo sapiens.
PN WO9404171-A.
PN 03-MAR-1994.
PF 11-AUG-1993; U07545.
PR 11-AUG-1992; US-925460.
PR 15-JUN-1993; US-925460.
PA (HARD) HARVARD COLLEGE.
PI Chicz RM, Hedley ML, Stern LJ, Strominger JL, Urban RG;
PI Vignali DA;
DR WPI; 94-082825/10.
PT Novel immunomodulatory peptide(s) and nucleic acids - useful for
PT treatment of autoimmune diseases, transplant rejection and for
PT vaccination
PS Disclosure; Page 48; 139pp; English.
CC The sequences given in R49291-505 and R46981-7038 represent peptide
CC fragments of naturally-occurring immunomodulatory proteins. These
CC fragments are between 10-30 residues in length and bind to a human
CC major histocompatibility complex (MHC) class II allotype. These
CC peptides may be used for therapy of autoimmune diseases, such as

Initial Score	=	4	Optimized Score	=	4	Significance	=	3.70
Residue Identity	=	33%	Matches	=	2	Mismatches	=	4
Gaps	=	0	Conservative Substitutions	=			=	0

```

> 0 <
0| 10
> 0 <

```

Results file sqlpir.res made by on Fri 19 May 95 8:47:13-PDT.

Results of the initial comparison of US-08-121-713B-1 (1-6) with:
a bank : PIR 43, all entries

*

*

[illegible]

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	6
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		

Scores:	Mean	Median	Standard Deviation
	1	3	1.03

Times:	CPU	Total Elapsed
	00:01:11.96	00:01:12.00

Number of residues:	22468834
Number of sequences searched:	75511
Number of scores above cutoff:	3753

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

The scores below are sorted by initial score. Significance is calculated based on initial score.

4 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
1. NEMSH	complement factor H precursor	1234	6	6	4.85 0
2. S47464	methyiaspartate mutase (EC 5.	483	6	6	4.85 0
3. S44573	hypothetical protein YK1609 -	259	6	6	4.85 0
4. S37984	probable purine nucleotide-bi	244	6	6	4.85 0

The list of other best scores is:

Sequence Name	Description	*** 2 standard deviations above mean ***	Init. Opt.	Sig. Frame
		Length Score	Score	
5. PS0249	porin - rice (strain Nihonbar	14	4	2.91 0
6. A0383	gene X protein - Escherichia	16	47350559	2.91 0
7. E60977	14-3-3 protein homolog, 15K -	19	4	2.91 0
8. S06487	ferredoxin, 15K, mitochondria	22	4	2.91 0
9. A54725	cecropin B - Chinese oak silk	35	4	2.91 0
10. S45125	hypothetical protein N2023 -	35	4	2.91 0
11. CKA0BP	cecropin B - Chinese oak silk	36	4	2.91 0
12. B31252	metallothionein II - yeast (C	37	4	2.91 0
13. A23617	conglutinin delta-2 small chain	37	4	2.91 0
14. B32037	ribosomal protein L36 - Bacil	37	4	2.91 0
15. R9BS36	ribosomal protein L36 - Bacil	37	4	2.91 0
16. R31V36	ribosomal protein L36 - liver	37	4	2.91 0
17. S11265	ribosomal protein V143 - yeast	39	4	2.91 0
18. B18891	licheninase (EC 3.2.1.73) II	40	4	2.91 0
19. A18891	licheninase (EC 3.2.1.73) I -	40	4	2.91 0
20. P00752	beta-fructofuranosidase (EC 3	40	4	2.91 0
21. E22565	R-phycoerythrin beta-3 chain	40	4	2.91 0
22. A24533	cuticle protein SC1 - flesh f	43	4	2.91 0
23. W5WLEP	E5 protein - European elk pap	43	4	2.91 0
24. BXSAL	antibacterial protein 1 - Sta	44	4	2.91 0
25. A28167	arylamine N-acetyltransferase	46	4	2.91 0
26. A60443	thymidylate synthase (EC 2.1.1	46	4	2.91 0
27. S10214	hexon-associated protein (VII	47	4	2.91 0
28. J00359	insulin - colubrid snake (Zao	51	4	2.91 0
29. INRS	insulin - western diamondback	51	4	2.91 0
30. INGS	insulin - goose	51	4	2.91 0
31. JC1197	metallothionein II - yeast (C	52	4	2.91 0
32. RUPF	rubredoxin - Pyrococcus furio	53	4	2.91 0
33. A05024	hypothetical protein 55 - liv	55	4	2.91 0
34. S01585	photosystem II protein psbK p	55	4	2.91 0
35. NTHB2	neurotoxin B-II - ribbon worm	55	4	2.91 0
36. NTHB4	neurotoxin B-IV - ribbon worm	55	4	2.91 0
37. TLEP54	tail fiber protein gp35 - pha	58	4	2.91 0
38. J00750	R21 protein - phage lambda	60	4	2.91 0
39. QXBP7L	hypothetical protein C-60 (ni	60	4	2.91 0
40. S25992	hypothetical protein 61 - liv	61	4	2.91 0
41. C34123	depressant toxin BJr2 - Esgp	61	4	2.91 0
42. D40667	neurotoxin - Naja naja	62	4	2.91 0
43. JK0221	short neurotoxin 1 - monocled	62	4	2.91 0
44. CKWK6	cecropin B precursor - cecrop	62	4	2.91 0

45. N1NJ1F	short neurotoxin 1 - Chinese	62	4	4	2.91	0
------------	------------------------------	----	---	---	------	---

1. US-08-121-713B-1 (1-6)

complement factor H precursor - mouse

ENTRY	NMESH	#type complete
TITLE	complement factor H precursor - mouse	
ALTERNATE_NAMES	protein beta-1-H	
ORGANISM	#formal name Mus musculus #common name house mouse	
DATE	30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 28-Oct-1994	
ACCESSIONS	A26154	
REFERENCE	A26154	
#authors	Kristensen, T.; Tack, B.F.	
#journal	Proc. Natl. Acad. Sci. U.S.A. (1986) 83:3963-3967	
#title	Murine protein H is comprised of 20 repeating units, 61 amino acids in length.	
	#cross-references PMID:86233353	
	#accession A26154	
	#molecule_type mRNA	
	#residues 1-1234 ##label KRI	
COMMENT	Two codominant alleles of factor H are present in mice.	
COMMENT	Factor H functions as a cofactor in the inactivation of C3b by serine proteinase I and also increases the rate of dissociation of the C3bBb complex (C3 convertase) and the (C3b)Nbb complex (C5 convertase) in the alternative complement pathway.	
GENETICS	#map position 1	
CLASSIFICATION	#superfamily complement factor H; complement factor H repeat homology	
KEYWORDS	complement; duplication; glycoprotein; plasma	
FEATURE		
1-18	#domain signal sequence #status predicted #label SIG\	
19-1234	#product complement factor H #label MP1\	
21-80	#domain complement factor H repeat homology #label FH01\	
85-141	#domain complement factor H repeat homology #label FH02\	
146-205	#domain complement factor H repeat homology #label FH03\	
210-262	#domain complement factor H repeat homology #label FH04\	
267-320	#domain complement factor H repeat homology #label FH05\	
325-384	#domain complement factor H repeat homology #label FH06\	
389-442	#domain complement factor H repeat homology #label FH07\	
448-505	#domain complement factor H repeat homology #label FH08\	
509-564	#domain complement factor H repeat homology #label FH09\	
569-622	#domain complement factor H repeat homology #label FH10\	
629-683	#domain complement factor H repeat homology #label FH11\	
690-743	#domain complement factor H repeat homology #label FH12\	
752-802	#domain complement factor H repeat homology #label FH13\	
808-861	#domain complement factor H repeat homology #label FH14\	
867-931	#domain complement factor H repeat homology #label FH15\	
936-989	#domain complement factor H repeat homology #label FH16\	
994-1048	#domain complement factor H repeat homology #label FH17\	
1053-1107	#domain complement factor H repeat homology #label FH18\	
1114-1168	#domain complement factor H repeat homology #label FH19\	
1172-1233	#domain complement factor H repeat homology #label FH20\	
21-66, 52-80, 85-129,		
114-141, 146-192,		
178-205, 210-251,		
237-262, 267-309,		
294-320, 325-374,		

357-385,389-431,
416-442,448-494,
477-505,509-553,
536-564,569-610,
597-622,629-672,
658-683,690-732,
718-743,752-791,
780-802,808-850,
836-861,867-920,
906-931,936-978,
964-989,994-1037,
1023-1048,
1053-1096,
1082-1107,
1114-1157,
1143-1168,
1172-1223,
1206-1233,
676,721,773,801,
1030,1061,1225

#disulfide_bonds #status predicted\

SUMMARY #length 1234 #molecular-weight 139081 #checksum 3676
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.85
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| | |
PLELFGVEVMCENGIWKPKCRDS
1090 X 1100 1110

2. US-08-121-713B-1 (1-6)
S47464 methylaspartate mutase (EC 5.4.99.1) - Clostridium

ENTRY S47464 #type complete
TITLE methylaspartate mutase (EC 5.4.99.1) - Clostridium
ORGANISM cochlearium
DATE #formal name Clostridium cochlearium
23-Nov-1994; #sequence_revision 23-Nov-1994; #text_change
23-Nov-1994
ACCESSIONS S47464
REFERENCE S47461
#authors Zelder, O.
#submission submitted to the EMBL Data Library, August 1994
#accession S47464
#status preliminary
#residues 1-483 ##label ZEL
#cross-references EMBL:X80997
SUMMARY #length 483 #molecular-weight 53574 #checksum 4915
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.85
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X

XCXNXI
| | |
DAYTRQNRVDECENGIKESEKAGRSI
100 X 110 120

3. US-08-121-713B-1 (1-6)
S44573 hypothetical protein YKL609 - yeast (Saccharomyces

ENTRY S44573 #type complete
TITLE hypothetical protein YKL609 - yeast (Saccharomyces
cerevisiae)
ORGANISM #formal name Saccharomyces cerevisiae
DATE 28-Sep-1994; #sequence_revision 28-Sep-1994; #text_change
28-Sep-1994
ACCESSIONS S44573
REFERENCE S44563
#authors Vandenbol, M.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger,

F.
Yeast (1994) 10:35-40
DNA sequencing of a 36.2 kb fragment located between the FAS1
and LAP4 loci of chromosome XI of Saccharomyces cerevisiae.

#journal
#title
#accession S44573
#status preliminary
#residues 1-259 ##label VAN
#cross-references EMBL:Z26877
SUMMARY #length 259 #molecular-weight 28620 #checksum 3202
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.85
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| | |
LVDILSITESCEGIDILIACNKSE
150 160 X 170

4. US-08-121-713B-1 (1-6)
S37984 probable purine nucleotide-binding protein YKL154w

ENTRY S37984 #type complete
TITLE probable purine nucleotide-binding protein YKL154w - yeast
(Saccharomyces cerevisiae)
ORGANISM #formal name Saccharomyces cerevisiae
DATE 03-May-1994; #sequence_revision 03-May-1994; #text_change
08-Dec-1994
ACCESSIONS S37984; S37796
REFERENCE S37976
#authors Vandenbol, M.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger,

F.
#submission submitted to the Protein Sequence Database, March 1994
#accession S37984
#molecule_type DNA
#residues 1-244 ##label VAN
#cross-references EMBL:Z28154
REFERENCE S37786
#authors Vandenbol, M.; Bolle, P.; Dion, C.; Portetelle, D.; Hilger,

F.
#submission submitted to the EMBL Data Library, September 1993
#description DNA sequencing of a 36.2 kb fragment located between the FAS1
and LAP4 loci of chromosome XI of *S. cerevisiae*.
#accession S37796
#molecule_type DNA
#residues 'TTLPLFISIIAHGIR', 1-244 ##label VA2
#cross-references EMBL:226877
GENETICS
#map_position 1L
KEYWORDS P-loop; purine nucleotide binding
FEATURE
45-52
51 #region purine nucleotide-binding motif A (P-loop)\
#binding site ATP/GTP (Lys) #status predicted
SUMMARY #length 244 #molecular-weight 26974 #checksum 2448
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.85
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| | |
LVDILSITSCENGIDILIACKNSE
140 X 150

5. US-08-121-713B-1 (1-6) porin - rice (strain Nihonbare) (fragment)

PS0249
ENTRY #type fragment
TITLE porin - rice (strain Nihonbare) (fragment)
ORGANISM #formal_name Oryza sativa #common_name rice
DATE 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
03-Feb-1994
ACCESSIONS PS0249
REFERENCE Tsugita, A.
#authors submitted to JIPID, April 1993
#contents callus
#accession PS0249
#molecule_type protein
#residues #length 14 #checksum 7764
SUMMARY
SEQUENCE

Initial Score = 4 Optimized Score = 4 Significance = 2.91
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| |
AVTFTDDHTANGIK
10 X

6. US-08-121-713B-1 (1-6) gene X protein - Escherichia coli plasmid R100 (fr

ENTRY A40383 #type fragment
TITLE gene X protein - Escherichia coli plasmid R100 (fragment)
ORGANISM #formal_name Escherichia coli
DATE 27-Nov-1991 #sequence_revision 27-Nov-1991 #text_change
18-Jun-1993
ACCESSIONS A40383
REFERENCE Inamoto, S.; Yoshioka, Y.; Ohtsubo, E.
#authors J. Biol. Chem. (1991) 266:10086-10092
#journal Site- and strand-specific nicking in vitro at oriT by the
#title TraY-TraI endonuclease of plasmid R100.
#cross-references MUID:91244768
#accession A40383
#status preliminary
#molecule_type DNA
#residues 1-16 ##label INA
#note sequence not compared to nucleotide translation

GENETICS
#genome plasmid
SUMMARY #length 16 #checksum 4
SEQUENCE

Initial Score = 4 Optimized Score = 4 Significance = 2.91
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| |
MKKWMLAICLIMFINGI
10 X

7. US-08-121-713B-1 (1-6) E60977 14-3-3 protein homolog, 15K - California sea hare

E60977 #type fragment
TITLE 14-3-3 protein homolog, 15K - California sea hare (fragment)
ORGANISM #formal_name Aplysia californica #common_name California sea
hare
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
03-Feb-1994
ACCESSIONS E60977
REFERENCE Sweatt, J.D.; Kennedy, T.E.; Wager-Smith, K.; Gawinowicz,
#authors M.A.; Barzilai, A.; Karl, K.A.; Kandel, E.R.
#journal Electrophoresis (1989) 10:152-157
#title Development of a database of amino acid sequences for
proteins identified and isolated on two-dimensional
polyacrylamide gels.

#accession E60977
#molecule_type protein
#residues 1-19 ##label SWE
SUMMARY #length 19 #checksum 4274
SEQUENCE

Initial Score = 4 Optimized Score = 4 Significance = 2.91
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

```
      X      X
      XCXNXI
      | |
ITKDYRKQIEKEINGI
      10  X  X

8. US-08-121-713B-1 (1-6)
   S06487 ferredoxin, 15K, mitochondrial - pig (fragment)

ENTRY      #type fragment
TITLE      ferredoxin, 15K, mitochondrial - pig (fragment)
ORGANISM   #formal name Sus scrofa domestica #common name domestic pig
DATE       30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
           26-May-1994

ACCESSIONS S06487
REFERENCE   S06485
            Driscoll, W.J.; Omdahl, J.L.
            Eur. J. Biochem. (1989) 185:181-187
            Characterization and N-terminal amino acid sequence of
            multiple ferredoxins in kidney and adrenal mitochondria.
            #cross-references M01D:90032674
            #accession S06487

            ##molecule_type protein
            ##residues_1-22 ##label DRI
CLASSIFICATION #superfamily glycine cleavage system protein H
KEYWORDS      electron transfer; mitochondrion
SUMMARY       #length 22 #checksum 9488
SEQUENCE
```

```
Initial Score = 4 Optimized Score = 4 Significance = 2.91
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
```

```
      X      X
      XCXNXI
      | |
SKFTDKVEWITTEGIGTVG
      10  X  20
```

```
9. US-08-121-713B-1 (1-6)
   A54725 cecropin B - Chinese oak silkmoth

ENTRY      #type complete
TITLE      cecropin B - Chinese oak silkmoth
ORGANISM   #formal name Antheraea pernyi #common_name Chinese oak
           silkmoth
DATE       11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change
           11-Nov-1994

ACCESSIONS A54725
REFERENCE   A54725
            Craig, A.G.; Engstroem, A.; Bennich, H.; Kamensky, I.
            Biomed. Environ. Mass Spectrom. (1987) 14:669-673
            Plasma desorption mass spectrometry coupled with conventional
            peptide sequencing techniques.
            #accession A54725
            ##status preliminary
            ##molecule_type protein
```

```
##residues_1-35 ##label CRA
KEYWORDS      amidation
SUMMARY       #length 35 #molecular-weight 3818 #checksum 6340
SEQUENCE

Initial Score = 4 Optimized Score = 4 Significance = 2.91
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
```

```
      X      X
      XCXNXI
      | |
IFKIEKVGRIIRNGIIKAGFAVAVL
      10  X  20
```

```
10. US-08-121-713B-1 (1-6)
   S45125 hypothetical protein N2023 - yeast (Saccharomyces
ENTRY      #type complete
TITLE      hypothetical protein N2023 - yeast (Saccharomyces cerevisiae)
ORGANISM   #formal name Saccharomyces cerevisiae
DATE       03-May-1994 #sequence_revision 02-Aug-1994 #text_change
           02-Aug-1994

ACCESSIONS S45125
REFERENCE   S45118
            Verhasselt, P.; Aert, R.; Voet, M.; Volckaert, G.
            submitted to the EMBL Data Library, January 1994.
            #submission Twelve open reading frames revealed on the 23.6 kbp segment
            #description flanking the centromere on the Saccharomyces cerevisiae
            chromosome XIV right arm.
            #accession S45125
            ##molecule_type DNA
            ##residues_1-35 ##label VER
            ##cross-references EMBL:X77395
```

```
GENETICS
#map_position 14
SUMMARY      #length 35 #molecular-weight 3966 #checksum 7961
SEQUENCE
```

```
Initial Score = 4 Optimized Score = 4 Significance = 2.91
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
```

```
      X      X
      XCXNXI
      | |
SIFCIPNIIAFFNNGIPTIISKIT
      20  X  X  30
```

```
11. US-08-121-713B-1 (1-6)
   CKA0BP cecropin B - Chinese oak silkmoth

ENTRY      #type complete
TITLE      cecropin B - Chinese oak silkmoth
ORGANISM   #formal name Antheraea pernyi #common_name Chinese oak
           silkmoth
DATE       17-Dec-1982 #sequence_revision 30-Sep-1988 #text_change
           30-Sep-1993
```

ACCESSIONS A01771
REFERENCE A91122
#authors Qu, X.; Steiner, H.; Engstrom, A.; Bennich, H.; Boman, H.G.
#journal Eur. J. Biochem. (1982) 127:219-224
#title Insect immunity: isolation and structure of cecropins B and D from pupae of the Chinese oak silk moth, *Antheraea pernyi*.
#cross-references MUID:83053368

#accession A01771
##molecule_type protein
##residues 1-36 ##label QUX
##note the carboxyl end is blocked
CLASSIFICATION #superfamily cecropin
SUMMARY #length 36 #molecular-weight 3905 #checksum 9328
SEQUENCE

Initial Score = 4 Optimized Score = 4 Significance = 2.91
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| |
IFKKIEKVGRIIRNGIIRAGPAVAVL
10 X 20

12. US-08-121-713B-1 (1-6)
B31252 metallothionein II - yeast (*Candida glabrata*) (fragment)

ENTRY B31252 #type fragments
TITLE metallothionein II - yeast (*Candida glabrata*) (fragments)
ORGANISM #formal name *Candida glabrata*
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993

ACCESSIONS B31252
REFERENCE A94212
#authors Mehra, R.K.; Tarbet, E.B.; Gray, W.R.; Winge, D.R.
#journal Proc. Natl. Acad. Sci. U.S.A. (1988) 85:8815-8819
#title Metal-specific synthesis of two metallothioneins and gamma-glutamyl peptides in *Candida glabrata*.

#cross-references MUID:89057829
#accession B31252
##molecule_type protein
##residues 1-37 ##label MEH
SUMMARY #length 37 #checksum 1778
SEQUENCE

Initial Score = 4 Optimized Score = 4 Significance = 2.91
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| |
CQYDCHCANCACENSASNECSQTCCK
10 X 20

13. US-08-121-713B-1 (1-6)
A23617 conglutin delta-2 small chain - narrow-leaved blue

ENTRY A23617 #type complete
TITLE conglutin delta-2 small chain - narrow-leaved blue lupine
ORGANISM #formal name *Lupinus angustifolius* #common_name narrow-leaved blue lupine
DATE 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 18-Jun-1993

ACCESSIONS A23617
REFERENCE A91358
#authors Lilley, G.G.; Inglis, A.S.
#journal FEBS Lett. (1986) 195:235-241
#title Amino acid sequence of conglutin delta, a sulfur-rich seed protein of *Lupinus angustifolius* L. Sequence homology with the C-III alpha-amylase inhibitor from wheat.

#accession A23617
##molecule_type protein
##residues 1-37 ##label LIL
SUMMARY #length 37 #molecular-weight 4598 #checksum 2819
SEQUENCE

Initial Score = 4 Optimized Score = 6 Significance = 2.91
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
| |
KQQLQQVNIHRCHENHIDRIQQQQEE
10 20 X 30

14. US-08-121-713B-1 (1-6)
B32307 ribosomal protein L36 - *Bacillus subtilis*

ENTRY B32307 #type complete
TITLE ribosomal protein L36 - *Bacillus subtilis*
ALTERNATE_NAMES ribosomal protein B
ORGANISM #formal name *Bacillus subtilis*
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1993

ACCESSIONS A32307; B32307
REFERENCE A32307
#authors Boylan, S.A.; Suh, J.W.; Thomas, S.M.; Price, C.W.
#journal J. Bacteriol. (1989) 171:2553-2562
#title Gene encoding the alpha core subunit of *Bacillus subtilis* RNA polymerase is cotranscribed with the genes for initiation factor 1 and ribosomal proteins B, S13, S11, and L17.

#cross-references MUID:89213940

#accession A32307
##molecule_type DNA
##residues 1-37 ##label BOY
##note sequence not compared to nucleotide translation
CLASSIFICATION #superfamily *Escherichia coli* ribosomal protein L36
KEYWORDS protein biosynthesis; ribosome
SUMMARY #length 37 #molecular-weight 4305 #checksum 3561
SEQUENCE

Initial Score = 4 Optimized Score = 4 Significance = 2.91
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNI
| |
VIRRGKVMVICENPKHKQKQ
20 X 30

```

15. US-08-121-713B-1 (1-6)
R5BS36      ribosomal protein L36 - Bacillus stearothermophilus

ENTRY       R5BS36      #type complete
TITLE       ribosomal protein L36 - Bacillus stearothermophilus
ALTERNATE_NAMES  ribosomal protein BL38; ribosomal protein II
ORGANISM    #formal_name Bacillus stearothermophilus
DATE        31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change
            30-Jun-1993

ACCESSIONS  S08566
REFERENCE   S07236
AUTHORS     Tanaka, I.; Kimura, M.; Kimura, J.; Dijk, J.
JOURNAL     FEBS Lett. (1984) 166:343-346
#title      The amino acid sequence of two small ribosomal proteins from
            Bacillus stearothermophilus.
#cross-references M01D:84108949
#accession   S08566
            #molecule_type protein
            #residues 1-37 #label TAN

CLASSIFICATION #superfamily Escherichia coli ribosomal protein L36
KEYWORDS       protein biosynthesis; ribosome
SUMMARY        #length 37 #molecular-weight 4361 #checksum 3946
SEQUENCE

```

Initial Score	=	4	Optimized Score	=	4	Significance	=	2.91
Residue Identity	=	33%	Matches	=	2	Mismatches	=	4
Gaps	=	0	Conservative Substitutions	=	0		=	0

X X
 XCXNXI
 | |
 VIRRRGKVMVICENPKK
 20 X 30
 > 0 <
 0 10
 > 0 <
 IntelliGenetic

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sqlspt.res made by on Fri 19 May 95 8:51:45-PDT.

Query sequence being compared:	US-08-121-713B-1 (1-6)
Number of sequences searched:	43470
Number of scores above cutoff:	3752

Results of the initial comparison of US-08-121-713B-1 (1-6) with:
Data bank : Swiss-Prot 31, all entries

100000-

Profession	Number of People (approx.)
Nurse	48,000
US Marine	45,000
Musician	40,000
Engineer	35,000
Researcher	30,000
Officer	25,000
Footballer	20,000
Scientist	15,000
Queen	10,000
University	5,000
Entrepreneur	5,000
Chef	5,000
Scientist	5,000

[illegible]

Cutoff score 0
Randomization group 0
Initial scores to save 45
Optimized scores to save 0

Alignments to save 15
Display context 10

SEARCH STATISTICS

Scores: Mean Median Standard Deviation
1 3 0.93
Times: CPU 00:00:43.02
Total Elapsed 00:00:43.00

Number of residues: 15335248
Number of sequences searched: 43470
Number of scores above cutoff: 3752

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

2 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Opt. Score	Sig. Frame
1. CFAH_MOUSE	COMPLEMENT FACTOR H PRECURSOR	1234	6	6 5.36 0
2. YK24_YEAST	HYPOTHETICAL 27.0 KD PROTEIN	244	6	6 5.36 0

The list of other best scores is:

Sequence Name	Description	Length	Init. Opt. Score	Sig. Frame
3. UKA4_HUMAN	UNKNOWN PROTEIN FROM 2D-PAGE	16	4	4 3.22 0
4. COXJ_ONCMY	CYTOKROME C OXIDASE POLYPEPT	23	4	4 3.22 0
5. CECB_ANTPE	CECROPIN B.	35	4	4 3.22 0
6. RL36_TETH	50S RIBOSOMAL PROTEIN L36 (RI	37	4	4 3.22 0
7. RL36_WARPO	50S RIBOSOMAL PROTEIN L36.	37	4	4 3.22 0
8. RL36_BACSU	50S RIBOSOMAL PROTEIN L36 (RI	37	4	4 3.22 0
9. RL36_BACST	50S RIBOSOMAL PROTEIN L36 (RI	37	4	4 3.22 0
10. CG25_LUPAN	CONGLUTIN DELTA-2 SMALL CHAIN	37	4	6 3.22 0
11. RL43_YEAST	60S RIBOSOMAL PROTEIN YL43 (F	39	4	4 3.22 0
12. ORN3_FLAOR	ORNATIN A3.	41	4	4 3.22 0
13. VES5_PAPVE	E5 PROTEIN.	43	4	4 3.22 0
14. CUPI_SARBU	LARVAL CUTICLE PROTEIN SC1 (F	43	4	4 3.22 0
15. GG11_STAHA	ANTIBACTERIAL PROTEIN 1 (GONO	44	4	4 3.22 0
16. HEX8_ADE40	HEXON-ASSOCIATED PROTEIN PREC	47	4	4 3.22 0
17. RB12_MOUSE	RAS-RELATED PROTEIN RAB-12 (F	49	4	4 3.22 0
18. MT2_CANGA	METALLOTHIONEIN-II.	51	4	4 3.22 0
19. INS_ZAODH	INSULIN.	51	4	4 3.22 0
20. INS_CROAT	INSULIN.	51	4	4 3.22 0
21. INS_ANGAN	INSULIN.	51	4	4 3.22 0

22. TCPT_ARATH	TRANSLATIONALLY CONTROLLED TU	53	4	4 3.22 0
23. RUBR_PYRFU	RUBREDOXIN (RD).	53	4	4 3.22 0
24. PSBK_WARPO	PHOTOSYSTEM II 4 KD REACTION	55	4	4 3.22 0
25. NXB4_CERLA	NEUROTOXIN B-IV.	55	4	4 3.22 0
26. NXB2_CERLA	NEUROTOXIN B-II.	55	4	4 3.22 0
27. Y60_LAMBD	HYPOTHETICAL NIN REGION PROTE	60	4	4 3.22 0
28. NXS1_NAJKA	SHORT NEUROTOXIN 1 (TOXIN C-6	62	4	4 3.22 0
29. NXS1_NAJAT	SHORT NEUROTOXIN 1 (COBROTOXI	62	4	4 3.22 0
30. CECB_HYACE	CECROPIN B PRECURSOR (IMMUNE-P	63	4	4 3.22 0
31. INAP_DROFU	MALE ACCESSORY GLAND SERINE-P	65	4	4 3.22 0
32. UREI_BACSB	UREASE ACCESSORY PROTEIN UREI	65	4	4 3.22 0
33. MALX_KLEPN	MALX PROTEIN (FRAGMENT).	66	4	4 3.22 0
34. XEN2_XENLA	XENOXIN-2.	66	4	4 3.22 0
35. RL21_HALMA	50S RIBOSOMAL PROTEIN HL21/HL	66	4	4 3.22 0
36. IGF2_CHICK	INSULIN-LIKE GROWTH FACTOR II	67	4	4 3.22 0
37. RNH_MYCSM	RIBONUCLEASE H (EC 3.1.26.4)	67	4	4 3.22 0
38. NUSG_CGDAB	TRANSCRIPTION ANTITERMINATION	70	4	4 3.22 0
39. VTB2_XENLA	VITELLOGENIN B2 PRECURSOR (VT	71	4	4 3.22 0
40. VTB1_XENLA	VITELLOGENIN B1 PRECURSOR (VT	71	4	4 3.22 0
41. HP11_ECTHA	HIGH POTENTIAL IRON-SULFUR PR	71	4	4 3.22 0
42. DISI_AKPI	DISINTEGRIN APPLAGIN (PLATELE	71	4	4 3.22 0
43. DISI_AKHA	DISINTEGRIN HALYSIN (PLATELET	71	4	4 3.22 0
44. DZ30_PEA	DISEASE RESISTANCE RESPONSE P	72	4	4 3.22 0
45. DISG_TRIGA	DISINTEGRIN TRIGRAMIN GAMMA (73	4	4 3.22 0

1. US-08-121-713B-1 (1-6)

CFAH_MOUSE COMPLEMENT FACTOR H PRECURSOR (PROTEIN BETA-1-H).

ID	CFAH_MOUSE	STANDARD;	PRT; 1234 AA.
AC	P06909;		
DT	01-JAN-1988 (REL. 06, CREATED)		
DT	01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)		
DT	01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)		
DE	COMPLEMENT FACTOR H PRECURSOR (PROTEIN BETA-1-H).		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;		
OC	EUTHERIA; RODENTIA.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RM	86233353		
RA	KRISTENSEN T., TACK B.F.;		
RL	PROC. NATL. ACAD. SCI. U.S.A. 83:3963-3967(1986).		
CC	-1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF		
CC	C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE		
CC	C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5		
CC	CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.		
CC	-1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN		
CC	MICE.		
CC	-1- SIMILARITY: CONTAINS 20 SUSHI (SCR) REPEATS.		
DR	EMBL; M12660; MMH.		
DR	PIR; A26154; NEMSH.		
DR	HSSP; P08603; 1HFT.		
KW	COMPLEMENT ALTERNATE PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI;		
FT	SIGNAL.		
FT	1 18		
FT	CHAIN 19 1234		
FT	DOMAIN 20 1224		
FT	REPEAT 20 81		
FT	REPEAT 84 142		

BY SIMILARITY.
COMPLEMENT FACTOR H.
20 X SUSHI (SCR) REPEATS.
SUSHI 1.
SUSHI 2.

FT	REPEAT	145	206	SUSHI 3.
FT	REPEAT	209	263	SUSHI 4.
FT	REPEAT	266	321	SUSHI 5.
FT	REPEAT	324	386	SUSHI 6.
FT	REPEAT	388	443	SUSHI 7.
FT	REPEAT	447	506	SUSHI 8.
FT	REPEAT	508	565	SUSHI 9.
FT	REPEAT	568	623	SUSHI 10.
FT	REPEAT	628	684	SUSHI 11.
FT	REPEAT	689	744	SUSHI 12.
FT	REPEAT	751	803	SUSHI 13.
FT	REPEAT	807	862	SUSHI 14.
FT	REPEAT	866	932	SUSHI 15.
FT	REPEAT	935	990	SUSHI 16.
FT	REPEAT	993	1049	SUSHI 17.
FT	REPEAT	1052	1108	SUSHI 18.
FT	REPEAT	1113	1169	SUSHI 19.
FT	REPEAT	1171	1234	SUSHI 20.
FT	DISULFID	21	66	BY SIMILARITY.
FT	DISULFID	52	80	BY SIMILARITY.
FT	DISULFID	85	129	BY SIMILARITY.
FT	DISULFID	114	141	BY SIMILARITY.
FT	DISULFID	146	192	BY SIMILARITY.
FT	DISULFID	178	205	BY SIMILARITY.
FT	DISULFID	210	251	BY SIMILARITY.
FT	DISULFID	237	282	BY SIMILARITY.
FT	DISULFID	267	309	BY SIMILARITY.
FT	DISULFID	294	320	BY SIMILARITY.
FT	DISULFID	325	374	BY SIMILARITY.
FT	DISULFID	357	385	BY SIMILARITY.
FT	DISULFID	389	431	BY SIMILARITY.
FT	DISULFID	416	442	BY SIMILARITY.
FT	DISULFID	448	494	BY SIMILARITY.
FT	DISULFID	477	505	BY SIMILARITY.
FT	DISULFID	509	553	BY SIMILARITY.
FT	DISULFID	536	564	BY SIMILARITY.
FT	DISULFID	569	610	BY SIMILARITY.
FT	DISULFID	597	622	BY SIMILARITY.
FT	DISULFID	629	672	BY SIMILARITY.
FT	DISULFID	658	683	BY SIMILARITY.
FT	DISULFID	690	732	BY SIMILARITY.
FT	DISULFID	718	743	BY SIMILARITY.
FT	DISULFID	752	791	BY SIMILARITY.
FT	DISULFID	780	802	BY SIMILARITY.
FT	DISULFID	808	850	BY SIMILARITY.
FT	DISULFID	836	861	BY SIMILARITY.
FT	DISULFID	867	920	BY SIMILARITY.
FT	DISULFID	906	931	BY SIMILARITY.
FT	DISULFID	936	978	BY SIMILARITY.
FT	DISULFID	964	989	BY SIMILARITY.
FT	DISULFID	994	1037	BY SIMILARITY.
FT	DISULFID	1023	1048	BY SIMILARITY.
FT	DISULFID	1053	1096	BY SIMILARITY.
FT	DISULFID	1082	1107	BY SIMILARITY.
FT	DISULFID	1114	1157	BY SIMILARITY.
FT	DISULFID	1143	1168	BY SIMILARITY.
FT	DISULFID	1172	1223	BY SIMILARITY.
FT	DISULFID	1206	1233	BY SIMILARITY.
FT	CARBOHYD	676	676	POTENTIAL.
FT	CARBOHYD	721	721	POTENTIAL.

FT	CARBOHYD	773	773	POTENTIAL.				
FT	CARBOHYD	801	801	POTENTIAL.				
FT	CARBOHYD	1030	1030	POTENTIAL.				
FT	CARBOHYD	1061	1061	POTENTIAL.				
FT	CARBOHYD	1225	1225	POTENTIAL.				
SQ	SEQUENCE	1234	1234	AA; 139082 MW; 8347878 CN;				
Initial Score	=	6	Optimized Score	=	6	Significance	=	5.36
Residue Identity	=	50%	Matches	=	3	Mismatches	=	3
Gaps	=	0	Conservative Substitutions	=	0	Gaps	=	0

X X
XCXNI
| | |
PLELFGQVEVMCENGWTEKPKCRDS
1090 X 1100 1110

2. US-08-121-713B-1 (1-6)
YKP4 YEAST HYPOTHETICAL 27.0 KD PROTEIN IN APE2-GPM1 INTERGEN

ID	YKP4_YEAST	STANDARD;	PRT;	244 AA.
AC	P36057;			
AD	01-JUN-1994	(REL. 29, CREATED)		
DT	01-JUN-1994	(REL. 29, LAST SEQUENCE UPDATE)		
DT	01-OCT-1994	(REL. 30, LAST ANNOTATION UPDATE)		
DT	01-OCT-1994	(REL. 30, LAST SEQUENCE UPDATE)		
DE	HYPOTHEICAL 27.0 KD PROTEIN IN APE2-GPM1 INTERGENIC REGION			
DE	YKL154W OR YKL609.			
GN	OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).			
OS	EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.			
OC	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=S288C;			
RC	94378720			
RA	VANDENBOL M., BOLLE P.-A., DION C., PORTETELLE D., HILGER F.;			
RL	YEAST 101:S35-S40(1994).			
CC	-1- SIMILARITY: DISTANTLY RELATED TO RAS SUPERFAMILY.			
DR	EMBL; Z26877; SCDCRH11.			
DR	EMBL; Z28154; SCYKL154W.			
DR	PIR; S37984; S37984.			
KW	HYPOTHEICAL PROTEIN; GTP-BINDING.			
FT	NP BIND 45 52 GIP (POTENTIAL).			
FT	NP BIND 87 91 GTP (POTENTIAL).			
FT	NP BIND 154 157 GTP (POTENTIAL).			
SQ	SEQUENCE 244 AA; 26974 MW; 308102 CN;			

Initial Score	=	6	Optimized Score	=	6	Significance	=	5.36
Residue Identity	=	50%	Matches	=	3	Mismatches	=	3
Gaps	=	0	Conservative Substitutions	=			=	0

X X
XCXNI
— — —
LVDILSITESENGIDILIAKNKSE
140 X 150

3. US-08-121-713B-1 (1-6)
UKA4 HUMAN UNKNOWN PROTEIN FROM 2D-PAGE OF EPIDERMAL KERATINO

ID UKA4 HUMAN STANDARD; PRT; 16 AA.
AC P31145;
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE UNKNOWN PROTEIN FROM 2D-PAGE OF EPIDERMAL KERATINOCYTES (SPOT 2120)
DE (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE.
RC TISSUE=KERATINOCYTES;
RM 93162043
RA RASMUSSEN H.H., VAN DAMME J., PUYPE M., GESSER B., CELIS J.E.,
RA VANDEKERCKHOVE J.;
RL ELECTROPHORESIS 13:960-969(1992).
CC -!- ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN PROTEIN IS: 6.74,
ITS MW IS: 21.2 KD.
DR AARHUS/GHENT-2DPAGE; 2120; IEF.
FT NON TER 1 1
FT UNSURE 4 5
FT UNSURE 7 7
FT NON CONS 9 10
FT NON TER 16 16
SQ SEQUENCE 16 AA; 1809 MW; 1385 CN;
Initial Score = 4 Optimized Score = 4 Significance = 3.22
Residue Identity = 66% Matches = 2 Mismatches = 1
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
|
NGIHLDNIFILN
X X 10

4. US-08-121-713B-1 (1-6)
COXJ_ONCMY CYTOCHROME C OXIDASE POLYPEPTIDE VIIA-LIVER (EC 1.

ID COXJ_ONCMY STANDARD; PRT; 23 AA.
AC P80333;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE CYTOCHROME C OXIDASE POLYPEPTIDE VIIA-LIVER (EC 1.9.3.1) (VIIC)
DE (FRAGMENT).
OS ONCORHYNCHUS MYKISS (RAINBOW TROUT) (SALMO GAIRDNERI).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;
OC OSTEICHTHYES; ACTINOPTERYGII; SALMONIFORMES.
RN [1]
RP SEQUENCE.
RC TISSUE=LIVER;
RM 94237150
RA FREUND R., KADENBACH B.;
RL EUR. J. BIOCHEM. 221:1111-1116(1994).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN
MITOCHONDRIAL ELECTRON TRANSPORT.

CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
DR PIR; S43632; S43632.
KW OXIDOREDUCTASE; INNER MEMBRANE; MITOCHONDRION.
FT NON TER 23 23
SQ SEQUENCE 23 AA; 2635 MW; 3126 CN;
Initial Score = 4 Optimized Score = 4 Significance = 3.22
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
|
NMVPEKQKLFQAXNGIPVHLF
10 X 20

5. US-08-121-713B-1 (1-6)
CECB_ANTPE CECROPIN B.

ID CECB_ANTPE STANDARD; PRT; 35 AA.
AC F01509;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DE CECROPIN B.
OS ANTHERA PEREYI (CHINESE OAK SILK MOTH).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; LEPIDOPTERA.
RN [1]
RP PARTIAL SEQUENCE.
RM 83053368
RA QU X.-M., STEINER H., ENGSTROM A., BENNICH H., BOMAN H.G.;
RL EUR. J. BIOCHEM. 127:219-224(1982).
RN [2]
RP SEQUENCE.
RM 88108273
RA CRAIG A.G., ENGSTROM A., BENNICH H., KAMENSKY I.;
RL BIOMED. ENVIRON. MASS SPECTROM. 14:669-673(1987).
CC -!- FUNCTION: CECROPINS HAVE LYTIC AND ANTIBACTERIAL ACTIVITY AGAINST
SEVERAL GRAM-POSITIVE & GRAM-NEGATIVE BACTERIA.

DR PIR; A01771; CKA0BP.
DR PIR; A54725; A54725.
DR PROSITE; PS00268; CECROPIN.
KW INSECT IMMUNITY; ANTIBIOTIC; HEMOLYMPH; AMIDATION; MULTIGENE FAMILY.
FT MOD RES 35 35
SQ SEQUENCE 35 AA; 3818 MW; 5337 CN;
Initial Score = 4 Optimized Score = 4 Significance = 3.22
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XCXNXI
|
IFKKIEKVGNIIRNGLIKAGPAVAVL
10 X 20

6. US-08-121-713B-1 (1-6)
 RL36_THETH 50S RIBOSOMAL PROTEIN L36 (RIBOSOMAL PROTEIN B).
 ID RL36 THETH STANDARD; PRT; 37 AA.
 AC P80256;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE 50S RIBOSOMAL PROTEIN L36 (RIBOSOMAL PROTEIN B).
 GN RPMJ OR REL36.
 OS THERMUS AQUATICUS (SUBSP. THERMOPHILUS).
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
 CC -!- SIMILARITY: BELONGS TO THE L36P FAMILY OF RIBOSOMAL PROTEINS.
 DR PROSITE; PS00826; RIBOSOMAL_L36.
 KW RIBOSOMAL PROTEIN.
 SQ SEQUENCE 37 AA; 4421 MW; 6411 CN;
 Initial Score = 4 Optimized Score = 4 Significance = 3.22
 Residue Identity = 33% Matches = 2 Mismatches = 4
 Gaps = 0 Conservative Substitutions = 0

X X
 XCXNXI
 | |
 VIRRHGRVVICENPKHKQKQKQ
 20 X 30

7. US-08-121-713B-1 (1-6)
 RL36_MARPO 50S RIBOSOMAL PROTEIN L36.
 ID RL36 MARPO STANDARD; PRT; 37 AA.
 AC P12142;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
 DE 50S RIBOSOMAL PROTEIN L36.
 GN RPL36 OR SECK.
 OS MARCHANTIA POLYMORPHA (LIVERWORT).
 OC CHLOROPLAST.
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; BRYOPHYTA; HEPATICOPSIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 89068687
 RA FUKUZAWA H., KOCHI T., SANO T., SHIRAI H., UMESONO K., INOKUCHI H.,
 RA OZEKI H., OHYAMA K.;
 RL J. MOL. BIOL. 203:333-351 (1988).
 RN [2]
 RP COMPLETE GENOME.
 RA OHYAMA K., FUKUZAWA H., KOCHI T., SHIRAI H., SANO T., SANO S.,
 RA UMESONO K., SHIRAI Y., TAKEUCHI M., CHANG Z., AOTA S., INOKUCHI H.,
 RA OZEKI H.;
 RL NATURE 322:572-574 (1986).
 CC -!- SIMILARITY: BELONGS TO THE L36P FAMILY OF RIBOSOMAL PROTEINS.

DR EMBL; X04465; CHMPXX.
 DR PIR; A05060; R5LV36.
 DR PROSITE; PS00828; RIBOSOMAL_L36.
 KW RIBOSOMAL PROTEIN; CHLOROPLAST.
 SQ SEQUENCE 37 AA; 4521 MW; 5888 CN;
 Initial Score = 4 Optimized Score = 4 Significance = 3.22
 Residue Identity = 33% Matches = 2 Mismatches = 4
 Gaps = 0 Conservative Substitutions = 0

X X
 XCXNXI
 | |
 MKIRASVRKICENCLIRRRRIMV
 10 X 20

8. US-08-121-713B-1 (1-6)
 RL36_BACSU 50S RIBOSOMAL PROTEIN L36 (RIBOSOMAL PROTEIN II) (BL38).
 ID RL36 BACSU STANDARD; PRT; 37 AA.
 AC P20278;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE 50S RIBOSOMAL PROTEIN L36 (RIBOSOMAL PROTEIN II) (RIBOSOMAL PROTEIN B)
 GN RPMJ.
 OS BACILLUS SUBTILIS.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 89213940
 RA BOYLAN S.A., SUH J.-W., THOMAS S.M., PRICE C.W.;
 RL J. BACTERIOL. 171:2553-2562 (1989).
 CC -!- SIMILARITY: BELONGS TO THE L36P FAMILY OF RIBOSOMAL PROTEINS.
 DR EMBL; M26414; M26414.
 DR PIR; B32307; B32307.
 DR SUBTILIST; BG11042; RPMJ.
 DR PROSITE; PS00828; RIBOSOMAL_L36.
 KW RIBOSOMAL PROTEIN.
 SQ SEQUENCE 37 AA; 4305 MW; 7120 CN;
 Initial Score = 4 Optimized Score = 4 Significance = 3.22
 Residue Identity = 33% Matches = 2 Mismatches = 4
 Gaps = 0 Conservative Substitutions = 0

X X
 XCXNXI
 | |
 VIRRKGVVICENPKHKQKQKQ
 20 X 30

9. US-08-121-713B-1 (1-6)
 RL36_BACST 50S RIBOSOMAL PROTEIN L36 (RIBOSOMAL PROTEIN II) (BL38).
 ID RL36 BACST STANDARD; PRT; 37 AA.
 AC P07841;
 DT 01-AUG-1988 (REL. 08, CREATED)

DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
 DE 50S RIBOSOMAL PROTEIN L36 (RIBOSOMAL PROTEIN II) (RIBOSOMAL PROTEIN B)
 DE (BL38).
 GN RPMJ.
 OS BACILLUS STEAROTHERMOPHILUS.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RP SEQUENCE.
 RM 84108949
 RA TANAKA I., KIMURA M., KIMURA J., DIJK J.;
 RL FEBS LETT. 166:343-346(1984).
 CC -!- SIMILARITY: BELONGS TO THE L36P FAMILY OF RIBOSOMAL PROTEINS.
 DR PIR; S08566; R5BS36.
 DR PROSITE; PS00828; RIBOSOMAL_L36.
 KW RIBOSOMAL PROTEIN.
 SQ SEQUENCE 37 AA; 4361 MW; 6570 CN;

Initial Score = 4 Optimized Score = 4 Significance = 3.22
 Residue Identity = 33% Matches = 2 Mismatches = 4
 Gaps = 0 Conservative Substitutions = 0

X X
 XCXNXI
 I I
 VIRRGKVMVICENPKHQKQK
 20 X 30

10. US-08-121-713B-1 (1-6)
 CG2S_LUPAN CONGLUTIN DELTA-2 SMALL CHAIN.

ID CG2S LUPAN STANDARD; PRT; 37 AA.
 AC P09930;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
 DE CONGLUTIN DELTA-2 SMALL CHAIN.
 OS LUPINUS ANGSTIFOLIUS (NARROW-LEAVED BLUE LUPINE).
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; FABALES;
 OC FABACEAE.
 RN SEQUENCE.
 RP SEQUENCE.
 RC STRAIN=CV. WHITE;
 RA LILLEY G.G., INGLIS A.S.;
 RL FEBS LETT. 195:235-241(1986).
 CC -!- SUBUNIT: DIMER OF A SMALL CHAIN AND A LARGE CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 DR PIR; A23617; A23617.
 KW SEED.
 FT DISULFID 8 8 INTERCHAIN (WITH C-29 OF LARGE CHAIN).
 FT DISULFID 20 20 INTERCHAIN (WITH C-17 OR C-18 OF LARGE
 FT CHAIN).
 FT DOMAIN 29 37 GLU/GLN-RICH.
 SQ SEQUENCE 37 AA; 4598 MW; 4779 CN;

Initial Score = 4 Optimized Score = 6 Significance = 3.22
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 XCXNXI
 I I
 KRQLOQVNLPHCHENHIDRIQQQQEE
 10 20 X 30

11. US-08-121-713B-1 (1-6)
 RL43_YEAST 60S RIBOSOMAL PROTEIN YL43 (FRAGMENT).

ID RL43 YEAST STANDARD; PRT; 39 AA.
 AC P05747;
 DT 01-NOV-1988 (REL. 09, CREATED)
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE 60S RIBOSOMAL PROTEIN YL43 (FRAGMENT).
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
 RN SEQUENCE.
 RP OKATA E., HIGO K.-I., ITOH T.;
 RA MOL. GEN. GENET. 195:544-546(1984).
 CC -!- SIMILARITY: BELONGS TO THE L29E FAMILY OF RIBOSOMAL PROTEINS.
 DR PIR; S11265; S11265.
 KW RIBOSOMAL PROTEIN.
 FT NON TER 39
 SQ SEQUENCE 39 AA; 4532 MW; 8733 CN;

Initial Score = 4 Optimized Score = 4 Significance = 3.22
 Residue Identity = 33% Matches = 2 Mismatches = 4
 Gaps = 0 Conservative Substitutions = 0

X X
 XCXNXI
 I I
 NHTAHNQTAKAHNGIKPKTKYKPS
 10 X 20 30

12. US-08-121-713B-1 (1-6)
 ORN3_PLAOR ORNATIN A3.

ID ORN3 PLAOR STANDARD; PRT; 41 AA.
 AC P25510;
 DT 01-MAY-1992 (REL. 22, CREATED)
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
 DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
 DE ORNATIN A3.
 OS PLACODELLA ORNATA (TURTLE LEECH).
 OC EUKARYOTA; METAZOA; ANNELIDA; HIRUDINEA.
 RN SEQUENCE.
 RP SEQUENCE.
 RM 92111479
 RA MAZUR P., HENZEL W.J., SEYMOUR J.L., LAZARUS R.A.;
 RL EUR. J. BIOCHEM. 202:1073-1082(1991).
 CC -!- FUNCTION: POTENT INHIBITOR OF FIBRINOGEN INTERACTION WITH PLATELET
 CC RECEPTORS EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. MAY PREVENT
 CC BLOOD FROM CLOTTING DURING EITHER FEEDING AND/OR STORAGE OF
 CC INGESTED BLOOD.
 CC -!- SIMILARITY: HIGH, TO THE OTHER P. ORNATA ORNATINS, AND TO M. DECORA

CC DECORSIN.
CC -!- SIMILARITY: SMALL, TO THE SNAKE FAMILY OF DISINTEGRINS.
DR PIR; S19621; S19621.
KW BLOOD COAGULATION; PLATELET; CELL ADHESION.
FT SITE 33 35
SQ SEQUENCE 41 AA; 4386 MW; 7138 CN;
Initial Score = 4 Optimized Score = 4 Significance = 3.22
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
XCXNXI
IKESGPNDKRCNGITCTVGKCKIG
10 X 20 X 30
13. US-08-121-713B-1 (1-6)
VE5_PAPVE E5 PROTEIN.
ID VE5_PAPVE STANDARD; PRT; 43 AA.
AC P11330;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 01-MAY-1991 (REL. 18, LAST ANNOTATION UPDATE)
DE E5 PROTEIN.
OS EUROPEAN ELK PAPILLOMAVIRUS (EEPV).
CC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RM 87219878
RA AHOIA H., BERGMAN P., STROEM A.C., MORENO-LOPEZ J., PETERSON U.;
RL GENE 50:195-205(1986).
DR EMBL; M15953; PAPPECG.
DR PIR; E29499; WSWLEP.
KW EARLY PROTEIN.
SQ SEQUENCE 43 AA; 5182 MW; 9855 CN;
Initial Score = 4 Optimized Score = 4 Significance = 3.22
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
XCXNXI
FLVWWDQFGRCENMQL
30 X 40 X
14. US-08-121-713B-1 (1-6)
CUP1_SARBU LARVAL CUTICLE PROTEIN SC1 (FRAGMENT).
ID CUP1_SARBU STANDARD; PRT; 43 AA.
AC P14485;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE LARVAL CUTICLE PROTEIN SC1 (FRAGMENT).
OS SARCOPHAGA BULLATA (FLESH FLY).

OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE.
RM 86037264
RA HENZEL W.J., MOLE J.E., MULLIGAN K., LIPKE H.;
RL J. MOL. EVOL. 22:39-45(1985).
CC -!- FUNCTION: COMPONENT OF THE CUTICLE OF THE LARVA OF FLESH FLY.
CC -!- SIMILARITY: WITH OTHER INSECT LARVAL CUTICLE PROTEINS.
DR PIR; A24553; A24553.
DR PROSITE; P500233; CUTICLE FLEXIBLE.
KW STRUCTURAL PROTEIN; CUTICLE. CUTICULAR *CONSENSUS SEQUENCE*.
FT SIMILAR 23 >43
FT NON_TER 43 43
SQ SEQUENCE 43 AA; 4556 MW; 9428 CN;
Initial Score = 4 Optimized Score = 4 Significance = 3.22
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
XCXNXI
ALETSNGIHFWAIGGDEH
X 10
15. US-08-121-713B-1 (1-6)
GG11_STAHA ANTIBACTERIAL PROTEIN 1 (GONOCOCCAL GROWTH INHIBIT
ID GG11_STAHA STANDARD; PRT; 44 AA.
AC P11697;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE ANTIBACTERIAL PROTEIN 1 (GONOCOCCAL GROWTH INHIBITOR 1).
OS STAPHYLOCOCCUS HAEMOLYTICUS.
OC PROKARYOTA; FIRMICUTES; COCCI; MICROCOCCACEAE.
RN [1]
RP SEQUENCE.
RM 88339821
RA WATSON D.C., YAGUCHI M., BISAILLON J.G., BEAUDET R., MOROSOLI R.;
RL BIOCHEM. J. 252:87-93(1988).
CC -!- THIS PROTEIN IS POSSIBLY THE SIGNAL SEQUENCE OF A SECRETED OR
CC MEMBRANE ASSOCIATED PROTEIN.
CC -!- SIMILARITY: HIGH, TO ANTIBACTERIAL PROTEINS 2 AND 3.
DR PIR; S00599; BXSAL.
KW ANTIBIOTIC; SIGNAL; FORMYLATION.
FT MOD_RES 1 1 FORMYLATION (POTENTIAL).
SQ SEQUENCE 44 AA; 4523 MW; 10878 CN;
Initial Score = 4 Optimized Score = 4 Significance = 3.22
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
XCXNXI
WGRMGTSIVGIVENGITVLGKIFGF
20 30 X 40

maryh@stic

stdin

NeWSprinter20

Fri May 19 10:53:29 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5

5. R07930 Tryptic fragment T22 (a) of r
6. R46721 Rabies virus major antigenic
7. R51694 B.thuringiensis serovar Japon
8. P10343 Somatostatin deriv. A57.
9. R41498 TNF inhibitory peptide X.
10. P10114 Somatostatin analogue.
11. R48995 Blastogenesis inducing peptid
12. R48238 S1033 from N.meningitidis IM2
13. R55149 Transglutaminase peptide.
14. R53567 Birch pollen major allergen *
15. P50732 Sequence of peptide having ep
16. R33359 Sequence of tryptic peptide d
17. R60258 Mouse NF-ATp DNA splice site.
18. R51825 Der f II derived peptide, DF
19. R36477 DFII-1(1-20), a Dermatophagoi
20. R27605 Endothelin antagonist #12 (Op
21. R27604 Endothelin antagonist #11 (Op
22. R27603 Endothelin antagonist #10 (Op
23. R27602 Endothelin antagonist #9 (Dpr
24. R12780 Saraphotoxin Ac-s6c.
25. R12779 Saraphotoxin 6Hse-s6c.
26. R07484 Blood pressure regulating pep
27. R07482 Blood pressure regulating pep
28. P60550 Fragment of UK rotavirus majo
29. R34153 CDR2 domain of human V beta 2
30. R28971 Selectin based anti-adhesion
31. R20801 uPA/GF domain-alpha1AT-P fusi
32. R20799 PRE1-uPA/GF domain fusion Jun
33. R36478 DFII-2(11-35), a Dermatophago
34. R51826 Der f II derived peptide, DF
35. R24867 Sequence of peptide fragment
36. P81308 Atrial natriuretic polypeptid
37. R32877 N-terminal of Rat alpha-amida
38. P20197 Sequence of [D-Trp22]-somatos
39. P93340 Amino terminus sequence of an
40. R15584 Immunopeptide #3 derived from
41. R15583 Immunopeptide #2 derived from
42. P81726 Sequence of novel serine prot
43. P80977 Sequence of novel serine prot
44. P81730 Sequence of novel serine prot
45. P81734 Sequence of novel serine prot

1. US-08-121-713B-2 (1-6)
R05222 Antigen GX5401FL encoded by Eimeria tenella genom
AC R05222 standard; protein; 2189 AA.
DT 02-AUG-1990 (first entry)
DE Antigen GX5401FL encoded by Eimeria tenella genomic DNA
KW Eimeria tenella; antigen GX5401FL; antigen GX5401; avian coccidiosis.
OS Eimeria tenella.
PN W09000403-A.
PD 25-JAN-1990.
PF 05-JUL-1989; U02918.
PR 05-JUL-1988; US-215162.
PA (GENE-) Genex Corp.
PI Anderson DM, McCandless RJ, Strausberg SL, Strausberg RL;
DR WPI; 90-051586/07.
DR N-PSDB; Q03324.

PT Cloned gene or fragment encoding antigenic protein -
PT which binds with antibodies against avian coccidia, and
PS Claim 10; Page 93; Fig 14; 134pp; English.
CC It is encoded by an open reading frame contained within the sequence of
CC clone 533 which was derived from an E. tenella genomic library screened
CC with radioactively labelled cDNA encoding the GX5401 antigen. It is of
CC about 250 Kd. It carries several repeated peptide sequences and
CC is rich in cysteine residues. The open reading frame also encodes a
CC potential signal sequence for protein secretion. Also new are an
CC expression vector contg. cloned gene, and host cells transformed with
CC the vector. The transformed cells are used in a vaccine to immunise
CC birds against avian coccidiosis. By labelling the peptides, they can be
CC used as a type-specific probe. May also be used in an assay to detect
CC Ab against the coccidia. The Abs are used to identify transformed cells
CC contg. the DNA.
SQ Sequence 2189 AA;
SQ 202 A; 40 R; 103 N; 148 D; 0 B; 236 C; 50 Q; 242 E; 0 Z; 317 G; 54 H;
SQ 47 I; 51 L; 66 K; 12 M; 55 F; 85 P; 165 S; 168 T; 17 W; 36 Y; 95 V;
Initial Score = 6 Optimized Score = 6 Significance = 5.47
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0
X X
CXNXIX
I I I
EASPCGNTHCLNTIGSYCECKDGY
440 X 450 460
2. US-08-121-713B-2 (1-6)
R37844 Sequence of tryptic peptide 2 of rat natural kille
ID R37844 standard; peptide; 8 AA.
AC R37844;
DT 01-OCT-1993 (first entry)
DE Sequence of tryptic peptide 2 of rat natural killer protease
DE (RNKP-1).
KW DNA fragmentation; fragmentin; lymphocyte serine granule protease;
KW apoptosis; therapy.
OS Rattus rattus.
PN W09311246-A.
PD 10-JUN-1993.
PF 25-NOV-1992; CA0515.
PR 25-NOV-1991; GB-024986.
PA (UYMA-) UNIV MANITOBA.
PI Greenberg AH;
DR WPI; 93-197064/24.
PT Lymphocyte serine granule protease and its DNA - effects DNA
PT fragmentation and induces apoptosis, used for treating tumours
PS Claim 6; Table 2, page 28; 57pp; English.
CC Fragmentin 2 is a lymphocyte serine granule protease having DNA
CC fragmenting and apoptosis-inducing activity. It has a non-reduced
CC apparent mol. wt. of 31 kD and a reduced apparent mol. wt. of 32 kD.
CC It is prepd. from rat natural killer tumour cell line RNK-16. Four
CC tryptic peptides of fragmentin 2 were prepared and sequenced. The AA
CC sequence is illustrated in Table 2 and compared to the deduced AA
CC sequences of three granule proteases; the rat natural killer cell
CC protease RNKP-1 which was cloned from the same RNK-16 tumour used to

CC purify fragmentin, the murine T cell granule proteases CCPI/Granzyme
CC B and Hanukah Factor (HF)/granzyme A. Fragmentin was highly
CC homologous to RNKP-1 differing in only two of forty-one
CC identifiable AAs. CCPI/Granzyme B was also closely related as it
CC differed in seven amino acids, while 21 amino acids did not match
CC the corresponding HF/Granzyme A sequence.

SQ Sequence 8 AA;
SQ 2 A; 0 R; 1 N; 1 D; 0 B; 1 C; 0 Q; 1 E; 0 Z; 1 G; 0 H;
SQ 1 I; 0 L; 0 K; 0 M; 0 F; 0 P; 0 S; 0 T; 0 W; 0 Y; 0 V;

Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 40% Matches = 2 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX
| |
ANEICAGD
X X

3. US-08-121-713B-2 (1-6) Sequence of tryptic peptide 2 of fragmentin 2.

R37843 standard; peptide; 8 AA.

ID R37843;
AC R37843;
DT 01-OCT-1993 (first entry)
DE Sequence of tryptic peptide 2 of fragmentin 2.
KW DNA fragmentation; fragmentin; lymphocyte serine granule protease;
KW apoptosis; therapy.
OS Rattus rattus.
PN W09311246-A.
PD 10-JUN-1993.
PF 25-NOV-1992; CA0515.
PR 25-NOV-1991; GB-024986.
PA (UYMA-) UNIV MANITOBA.
PI Greenberg AH;
DR WPI; 93-197064/24.

PT Lymphocyte serine granule protease and its DNA - effects DNA
PT fragmentation and induces apoptosis, used for treating tumours
PS Claim 6; Table 2, page 28; 57pp; English.
CC Fragmentin 2 is a lymphocyte serine granule protease having DNA
CC fragmenting and apoptosis-inducing activity. It has a non-reduced
CC apparent mol. wt. of 31 kD and a reduced apparent mol. wt. of 32 kD.
CC it is prep'd. from rat natural killer tumour cell line RNK-16. Four
CC tryptic peptides of fragmentin 2 were prepared and sequenced. The AA
CC sequence is illustrated in Table 2 and compared to the deduced AA
CC sequences of three granule proteases; the rat natural killer cell
CC protease RNKP-1 which was cloned from the same RNK-16 tumour used to
CC purify fragmentin, the murine T cell granule proteases CCPI/Granzyme
CC B and Hanukah Factor (HF)/granzyme A. Fragmentin was highly
CC homologous to RNKP-1 differing in only two of forty-one
CC identifiable AAs. CCPI/Granzyme B was also closely related as it
CC differed in seven amino acids, while 21 amino acids did not match
CC the corresponding HF/Granzyme A sequence.

SQ Sequence 8 AA;
SQ 2 A; 0 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 1 E; 0 Z; 1 G; 0 H;
SQ 1 I; 0 L; 0 K; 0 M; 0 F; 0 P; 0 S; 0 T; 0 W; 0 Y; 0 V;
SQ 1 Others;

Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 60% Matches = 3 Mismatches = 2
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX
| |
ANEIXAGD
X X

4. US-08-121-713B-2 (1-6) Hepatitis C virus (HCV) epitope.

R35986 Hepatitis C virus (HCV) epitope.

ID R35986 standard; protein; 8 AA.
AC R35986;
DT 24-MAY-1993 (first entry)
DE Hepatitis C virus (HCV) epitope.
KW Hepatitis; liver disease; HCV; monoclonal antibody; epitope;
KW immobilised reagent; immunoassay; diagnosis; detection; treatment;
KW infection.
OS Hepatitis C virus type 1.
PN W09300365-A.
PD 07-JAN-1993.
PF 24-JUN-1992; U05388.
PR 24-JUN-1991; US-722489.
PA (CHIR) CHIRON CORP.
PI Chien DY, Rutter W;
DR WPI; 93-036334/04.
PT Polypeptide(s) comprising truncated hepatitis C virus sequences -
PT for detection, prevention and treatment of hepatitis C infection
PS Example A; Page 37; 80pp; English.
CC This octamer was found to be immunoreactive with anti-HCV anti-sera.
CC In the epitope mapping experiment three different samples of anti-sera
CC were reacted with the peptide octamer, and then incubated with
CC HRP-labelled goat anti-human Ig antisera, to enable detection of
CC binding. This epitope starts from amino acid 1762 of the HCV
CC polyprotein.

SQ Sequence 8 AA;
SQ 1 A; 0 R; 1 N; 0 D; 0 B; 0 C; 0 Q; 1 E; 0 Z; 0 G; 0 H;
SQ 1 I; 1 L; 1 K; 1 M; 0 F; 0 P; 0 S; 0 T; 1 W; 0 Y; 0 V;

Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 40% Matches = 2 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX
| |
AKLMWNEI
X X

5. US-08-121-713B-2 (1-6) Tryptic fragment T22 (a) of rat phospholipase A2 i

R07930 Tryptic fragment T22 (a) of rat phospholipase A2 i

ID R07930 standard; protein; 9 AA.
AC R07930;
DT 14-JAN-1991 (first entry)
DE Tryptic fragment T22 (a) of rat phospholipase A2 inhibitor.

KW Human N-lipoic acid; placenta; inflammation reduction; arthritis;
KW rat phospholipase A2 inhibitor; tryptic fragment T22 (a).
OS Homo sapiens.
PN US4950646-A.
PD 21-AUG-1990. 929199.
PF 10-JAN-1986; US-690146.
PR 10-JAN-1985; US-690146.
PR 15-MAR-1985; US-712376.
PR 14-AUG-1985; US-765877.
PR 05-SEP-1985; US-772892.
PR 10-JAN-1986; US-929199.
PA (BIOJ) BIOGEN NV.
PI Wallner BP, Pepinsky RB, Garwin JL, Schindler DG, Huang KS;
DR WPI; 90-274549/36.
PT Pure fragment of human lipoic acid - useful for reducing
PT inflammation or for treating arthritis, etc.
PS Disclosure; Fig 2; 51pp; English.
CC Rat phospholipase A2 inhibitor protein was isolated from the
CC extracellular supernatant of rat peritoneal exudate cells.
CC For tryptic fragment T22 (b) see R07931.
CC The protein was isolated and sequenced to produce oligonucleotide
CC probes in order to identify human lipoic acid and N-lipoic acid.
CC See also Q05805-25, Q06581, R07926-37 and R07956-66.
SQ Sequence 9 AA;
SQ 0 A; 0 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 2 E; 0 Z; 0 G; 0 H;
SQ 2 I; 0 L; 1 K; 1 M; 0 F; 0 P; 1 S; 0 T; 0 W; 0 Y; 0 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CXNIX
I I
SEIDWNEIK
X X

6. US-08-121-713B-2 (1-6)
R46721 Rabies virus major antigenic site III R333E mutant
ID R46721 standard; protein; 11 AA.
AC R46721;
DT 26-SEP-1994 (first entry)
DE Rabies virus major antigenic site III R333E mutant.
KW Rabies virus; avirulent; double mutant; SAD Berne strain; SAG2;
KW live vaccine; Control Virus Standard strain; CVS; glycoprotein G505;
KW antigen; immunogen.
OS Rabies Virus.
FH Key Location/Qualifiers
FT Peptide 1..11
FT /label= antigenic site III
FT /note= "amino acids 330-340 of the CVS strain
FT (identical to site III from SAD Berne
FT strain)"
FT Misc difference 4
FT /note= "wild-type Arg333 is substd. by Glu"
PN FR2693655-A.
PD 21-JAN-1994.
PF 20-JUL-1992; 008947.

PR 20-JUL-1992; FR-008947.
PA (VIRB-) VIRBAC SA.
PI Benejean J, Coulon P, Flamand A, Lefay F, Tuffereau M;
DR WPI; 94-058963/08.
PT Avirulent anti-rabies vaccine with double mutation in position
PT 333 of glyco-protein - of SAD strain, esp. having glutamic acid
PT instead of Arginine, stable against reversion and with good
PT protective activity
PS Claim 6; 16pp; French.
CC An anti-rabies vaccine comprises a double mutant form of the external
CC glycoprotein. In the mutant strain the codon at position 333 differs
CC by at least 2 nucleotides from the wild-type Arg codon. The pref.
CC mutation (specifically claimed) is from Arg to Glu. The mutation is
CC stable to reversion and results in an attenuated virus for use in
CC live vaccines. The specification does not include any sequence
CC listings; R46721 is major antigenic site III (i.e. amino acids 330-
CC 340) of the CVS strain (see Geneseq P40082 for the full-length
CC glycoprotein sequence) modified as described in the claims.
SQ Sequence 11 AA;
SQ 0 A; 0 R; 1 N; 0 D; 0 B; 0 C; 0 Q; 2 E; 0 Z; 0 G; 0 H;
SQ 2 I; 0 L; 1 K; 0 M; 0 F; 1 P; 1 S; 1 T; 1 W; 0 Y; 1 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CXNIX
I I
KSVETWNEIIP
X 10

7. US-08-121-713B-2 (1-6)
R51694 B.thuringiensis serovar Japonensis insecticidal pr
ID R51694 standard; peptide; 13 AA.
AC R51694;
DT 04-NOV-1994 (first entry)
DE B.thuringiensis serovar Japonensis insecticidal protein N-terminal.
KW Insecticidal protein; Coleoptera larvae; Bulbui fungus;
KW Bacillus thuringiensis; serovar Japonensis.
OS Bacillus thuringiensis (serovar Japonensis, strain Bulbui).
FH Key Location/Qualifiers
FT Misc difference 1..2
FT /note= "undefined"
PN J06065292-A.
PD 08-MAR-1994.
PF 11-AUG-1992; 213886.
PR 11-AUG-1992; JP-213886.
PA (KUBI) KUBOTA CORP.
DR WPI; 94-121220/15.
PT Insecticidal protein and DNA from Bacillus thuringiensis serovar
PT Japonensis strain Bulbui - useful in insecticides against
PT Coleoptera insects
PS Disclosure; Page 4; 18pp; Japanese.
CC An insecticidal protein with activity against Coleoptera
CC insect larvae has been isolated from Bacillus thuringiensis serovar
CC Japonensis strain Bulbui. The N-terminal amino acid sequence was
CC determined and used for the design of a probe to clone DNA coding

CC for the insecticidal protein.
SQ Sequence 13 AA;
SQ 1 A; 0 R; 3 N; 1 D; 0 B; 0 C; 1 Q; 1 E; 0 Z; 0 G; 0 H;
SQ 2 I; 1 L; 0 K; 0 M; 0 F; 1 P; 0 S; 0 T; 0 W; 0 Y; 0 V;
SQ 2 Others;

Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX
| |
XXPNQNEIIDAL
X 10

8. US-08-121-713B-2 (1-6)
P10343 Somatostatin deriv. A57.

ID P10343 standard; Protein; 13 AA.
AC P10343; 1992 (first entry)
DT 15-DEC-1992 (first entry)
DE Somatostatin deriv. A57.
KW Growth Hormone secretion; GH; diabetes mellitus; angiopathy;
KW acromegaly; diagnosis.
OS Synthetic.

FH Key Location/Qualifiers
FT Modified site 1
FT /note= "(4Cl)Phe"
FT Modified site 3
FT /label= Nle
FT Misc difference 13
FT /note= "Cys-NH-CH(CH₂OH)2,
FT Cys-NH-CH(CH₂OH)CH₂CH₂OH,
FT Cys-NH-CH(CH₂OH)CH₂CH₂CH₂OH or
FT as amide formed with 4-amino-pentanoic
FT acid lactone"
FN CH-621770-A.
PD 27-FEB-1981.

PF 11-SEP-1980; 125375.
PR 23-FEB-1976; CH-002175.
PA (SANO) SANDOZ AG.
PI Sandrin E. Bauer W;
DR WPI; 81-21515D/13 (21515D).
PT Somatostatin derivs. prodrn. - useful for treating diabetes,
PT acromegalia and angiopathy
PS Example 1; Page 7; 8pp; German.
CC This peptide is an example of a generic formula for somatostatin
CC derivs. which inhibit secretion of growth hormone and are useful to
CC treat diabetes mellitus, acromegaly, angiopathy and in diagnosis.
CC See P10308-P10348.

SQ Sequence 13 AA;
SQ 0 A; 0 R; 1 N; 0 D; 0 B; 2 C; 0 Q; 0 E; 0 Z; 0 G; 0 H;
SQ 0 I; 0 L; 1 K; 0 M; 4 F; 0 P; 1 S; 2 T; 1 W; 0 Y; 0 V;
SQ 1 Others;

Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX
| |
FCXNFFWKTFTSC
X X 10

9. US-08-121-713B-2 (1-6)
R41498 TNF inhibitory peptide X.

ID R41498 standard; peptide; 14 AA.
AC R41498;
DT 23-FEB-1994 (first entry)
DE TNF inhibitory peptide X.
KW Tumour necrosis factor; TNF; inhibition; solid phase synthesis; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT Disulfide bond 3..12
FT /note= "Optional di-sulphide bond"

PN J05194594-A.
PD 03-AUG-1993.
PF 21-JAN-1992; 029044.
PR 21-JAN-1992; JP-029044.
PA (SAGA) SAGAMI CHEM RES CENTRE.
DR WPI; 93-282916/36.
PT TNF inhibitory novel peptide(s) - include N-terminal amino Gp.
PT which is opt. modified with acetyl, T-butoxy-carbonyl or
PT benzyl-oxy-carbonyl Gp. and C-terminal carboxy Gp. is opt.
PT amidated
PS Claim 1; Page 6; 8pp; Japanese.

CC The sequences given in R41489-99 are tumour necrosis factor (TNF)
CC inhibitory peptides. They may optionally be modified at the N-
CC terminal with an acetyl, t-butoxycarbonyl or benzylloxycarbonyl, and
CC at the C-terminal they are optionally amidated. These peptides are
CC produced by solid phase synthesis methods and may be produced at low
CC cost.

SQ Sequence 14 AA;
SQ 0 A; 0 R; 1 N; 0 D; 0 B; 2 C; 1 Q; 1 E; 0 Z; 1 G; 1 H;
SQ 0 I; 3 L; 0 K; 0 M; 0 F; 0 P; 2 S; 1 T; 0 W; 0 Y; 1 V;

Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX
| |
SLCINGTVHLSQCE
X 10

10. US-08-121-713B-2 (1-6)
P10114 Somatostatin analogue.

ID P10114 standard; peptide; 14 AA.
AC P10114;
DT 15-DEC-1992 (first entry)
DE Somatostatin analogue.
KW Deprotection; selenocysteine; Sec.

OS Synthetic. Location/Qualifiers
FH Key Modified site 3
FT /label= Cys, OTHER
FT /note= "selenocysteine (Sec)"
FT Modified site 14
FT /label= Cys, OTHER
FT /note= "selenocysteine (Sec)"
FT Disulfide bond 3..14
FT Misc difference 8
FT /note= "D- or L- form residue"
PN DD-146289-A.
PD 04-FEB-1981.
PF 27-SEP-1979; 215859.
PR 27-SEP-1979; DD-215859.
PA (HART/) HARTRODT B.
PI Hartrodt B, Neubert K, Jakubke HD, Krabiell U;
DR WPI; 81-30862D/18 (30862D).
PT Somatostatin and analogues prodn. from oligopeptide fragments -
PT useful for treating acromegalia, angiopathy and diabetes
PS Claim 1; Page 33; 36pp; German.
CC The prodn. of the somatostatin analogue having the formula given
CC below is claimed. The process is relatively simple, esp. it avoids
CC acidolytic deprotection of Trp-8 (needed in known processes) which
CC can cause side reactions. The oligopeptide starting materials can
CC be made by the active ester method which requires only minimal side-
CC chain protection. The method is esp. applied to produce a cpd.
CC where amino acids 3 and 14 are Sec and Trp-8 is in D-form, which
CC cannot be made by other methods. The prod. is useful in treating
CC acromegalia, angiopathy and diabetes mellitus, and as diagnostic
CC agents.
SQ Sequence 14 AA;
SQ 1 A; 0 R; 1 N; 0 D; 0 B; 2 C; 0 Q; 0 E; 0 Z; 1 G; 0 H;
SQ 0 I; 1 L; 1 K; 0 M; 3 F; 0 P; 1 S; 2 T; 1 W; 0 Y; 0 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
CXNXX
I I
AGCLNFFWKTFTSC
X 10

11. US-08-121-713B-2 (1-6)
R48995 Blastogenesis inducing peptide #1.

ID R48995 standard; peptide; 15 AA.
AC R48995;
DT 14-SEP-1994 (first entry)
DE Blastogenesis inducing peptide #1.
KW Blastogenesis; mite-sensitised; lymphocyte; homodimer; heterodimer;
KW solid phase synthesis; diagnosis; mite allergy; disulphide bond.
OS Synthetic.
PN WO9404572-A.
PD 03-MAR-1994.
PF 10-AUG-1993; J01127.
PR 14-AUG-1992; JP-216955.

PA (ASAK) ASahi BREWERIES LTD.
PA (TORI) TORII & CO LTD.
PI Ando T, Ikeda S, Okumura Y;
DR WPI; 94-083115/10.
PT Peptides which cause blastogenesis of mite-sensitised lymphocytes
PT -as diagnostic agents for mite allergy.
PS Claim 2; Page 26; 33pp; Japanese.
CC The sequences given in R48995-96 are peptides which cause
CC blastogenesis in mite-sensitised lymphocytes. These peptides may be
CC used as individual peptides, homo- or heterodimers. These peptides
CC were produced by standard methods of solid phase synthesis and may
CC be used as diagnostic agents for the diagnosis of mite allergy. The
CC dimeric peptides are formed by disulphide bonds between Cys8 or Cys5
CC respectively.
SQ Sequence 15 AA;
SQ 1 A; 0 R; 2 N; 3 D; 0 B; 1 C; 1 Q; 1 E; 0 Z; 0 G; 0 H;
SQ 1 I; 0 L; 3 K; 0 M; 0 F; 0 P; 0 S; 0 T; 0 W; 0 Y; 2 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
CXNXX
I I
DQDVKDCANNEIKK
I 10 X
12. US-08-121-713B-2 (1-6)
R48238 S1033 from N.meningitidis IM2169 transferrin recep
ID R48238 standard; peptide; 17 AA.
AC R48238;
DT 15-JUL-1994 (first entry)
DE S1033 from N.meningitidis IM2169 transferrin receptor Tbp1 subunit.
KW Transferrin receptor; Tbp1 subunit; strain IM2169; meningitis vaccine;
KW Iron transport protein; iron chelator; internal tryptic peptide.
OS Neisseria meningitidis (strain IM2169).
FH Key Location/Qualifiers
FT Misc difference 16
FT /note= "undetermined amino acid"
FT Misc difference 17
FT /note= "undetermined amino acid"
PN FR2692592-A.
PD 24-DEC-1993.
PF 19-JUN-1992; 007493.
PR 19-JUN-1992; FR-007493.
PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS.
PA (TRGE) TRANSGENE SA.
PI Bloch M, Bouchon-Theisen B, Jacobs E, Legrain M;
PI Mazarin V, Shryvers AB, Schryvers AB;
DR WPI; 94-028254/04.
PT DNA coding for Neisseria meningitidis proteins - namely
PT transferrin receptor subunits
PS Example 2; Page 22; 60pp; French.
CC The transferrin receptor was isolated from a lysate of Neisseria
CC meningitidis strain IM2169. The subunits Tbp1 and Tbp2 were separated
CC and subjected to microsequencing at the N-terminal end. The Tbp1
CC subunit was subjected to trypsin digestion in order to obtain

CC internal peptides. See R48236-R48239 for tryptic peptide sequences.
SQ Sequence 17 AA;
SQ 2 A; 0 R; 2 N; 0 D; 0 B; 0 C; 0 Q; 3 E; 0 Z; 2 G; 0 H;
SQ 2 I; 0 L; 0 K; 0 M; 0 F; 0 P; 2 S; 1 T; 0 W; 1 Y; 0 V;
SQ 2 Others;
Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
CXNXIX
I I
TAGSSGAINEIYENXX
X 10 X
13. US-08-121-713B-2 (1-6)
R55149 Transglutaminase peptide.
ID R55149 standard; peptide; 18 AA.
AC R55149;
DT 16-DEC-1994 (first entry)
DE Transglutaminase peptide.
KW Pagrus major; enzyme; acyl transfer; gamma-carboxamide;
KW glutamine.
OS Pagrus major.
PN J06113844-A.
PD 26-APR-1994.
PF 15-DEC-1992; 334224.
PR 19-AUG-1992; JP-220296.
PA (AJIN) AJINOMOTO KK.
DR WPI; 94-172742/21.
PT Compsn. contg. trans-glutaminase derived from fish liver - useful
PT for prepn. of protein-gelled prod., esp. food prod.
PS Disclosure; Page 11; 26pp; Japanese.
CC A compen. contains at least 0.5 units per mg protein of a novel
CC transglutaminase (TGase). The TGase is derived from fish liver
CC (esp. Teragra chalcogramma and Pagrus major) and catalyses the
CC acyl transfer reaction between the gamma-carboxamide gp. of the
CC glutamine residue of a peptide chain and various prim. amines.
CC TGase is useful for prepn. of protein-gelled prods., esp. food.
SQ Sequence 18 AA;
SQ 2 A; 3 R; 1 N; 0 D; 0 B; 0 C; 1 Q; 3 E; 0 Z; 2 G; 0 H;
SQ 1 I; 1 L; 0 K; 0 M; 0 F; 1 P; 1 S; 1 T; 0 W; 0 Y; 1 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
CXNXIX
I I
AGRRVTSPSNEIAEQRL
10 X
14. US-08-121-713B-2 (1-6)
R53567 Birch pollen major allergen "Bet v I" amino acids

ID R53567 standard; peptide; 18 AA.
AC R53567;
DT 29-NOV-1994 (first entry)
DE Birch pollen major allergen "Bet v I" amino acids 93-110.
KW Order Fagales; tree pollen; birch; hazel; alder; major allergen;
KW allergy; T-cell epitope; Bet v I; Cor a I; Aln g I; tolerance.
OS Synthetic.
PN W09410194-A.
PD 11-MAY-1994.
PF 25-OCT-1993; AT0163.
PR 27-OCT-1992; AT-002125.
PR 14-JAN-1993; AT-000043.
PA (BIOM-) BIOMAY PRODN & HANDELSGES MBH.
PI Breitenbach M, Ebner C, Ferreira F, Kraft D, Rumpold H;
PI Scheiner O, Schenk S, Szepfalusi Z, Valenta R;
DR WPI; 94-167383/20.
DT New peptide(s) derived from tree pollen allergens - able to
PT induce T cell tolerance, useful in diagnoses and therapy of
PT allergies
PS Claim 1; Page 10; 12pp; German.
CC The 17kD major allergens from trees of the Order Fagales (esp.
CC birches, hazels and alders) are highly homologous. Peptides derived
CC from the T-cell epitopes of these allergens, partic. from the birch
CC Bet v I allergen are useful for diagnosing tree pollen allergy and
CC for stimulating or blocking T-cells of allergic patients in an
CC allergen-specific manner. The peptides can also be used to provoke
CC tolerance to the allergen-specific T-cells. The peptides have one
CC of the sequences R53560-R53569.
SQ Sequence 18 AA;
SQ 1 A; 0 R; 1 N; 2 D; 0 B; 0 C; 0 Q; 2 E; 0 Z; 1 G; 0 H;
SQ 3 I; 1 L; 2 K; 0 M; 0 F; 1 P; 1 S; 2 T; 0 W; 0 Y; 1 V;
Initial Score = 4 Optimized Score = 4 Significance = 3.65
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
CXNXIX
I I
DTLEKISNEIKIVATPDG
X 10
15. US-08-121-713B-2 (1-6)
P50732 Sequence of peptide having epidermal growth factor
ID P50732 standard; peptide; 18 AA.
AC P50732;
DT 29-NOV-1991 (first entry)
DE Sequence of peptide having epidermal growth factor (EGF)-like
DE activity.
KW De-fleecing agent; autophosphorylation inducer; antiulcer agent.
PN W08501284-A.
PD 28-MAR-1985.
PF 14-SEP-1984; U01459.
PR 14-SEP-1983; IL-069719.
PA (KOMORI) KOMORIYA A.
PI Komoriya A, Meyers CA, Schlessinger J;
DR WPI; 85-087027/14.

PT Polypeptide(s) with epidermal growth factor-like activity - also
 PR useful for de-fleecing sheep
 PS Claim 6; Page 10; 16pp; English.
 CC The peptides of the invention contain a 10 AA residue SQ of the
 CC natural EGF-like hormone corresp. to the SQ between Cys(20) and
 CC Cys(31). They activate EGF-sensitive kinase and induce
 CC autophosphorylation of the EGF receptors and other endogenous
 CC membrane proteins. They also stimulate thymidine incorporation into
 CC human foreskin fibroblasts. Dose is sufficient to give at least
 CC 50 micro m blood concn.
 SQ Sequence 18 AA;
 SQ 0 A; 0 R; 1 N; 1 D; 0 B; 3 C; 0 Q; 1 E; 0 Z; 2 G; 1 H;
 SQ 1 I; 2 L; 0 K; 1 M; 0 F; 0 P; 2 S; 1 T; 0 W; 1 Y; 1 V;
 Initial Score = 4 Optimized Score = 4 Significance = 3.65
 Residue Identity = 33% Matches = 2 Mismatches = 4.4
 Gaps = 0 Conservative Substitutions = 0

X X
 CXNX
 | |
 CLNGVCMHIESLDSY
 X X 10
 > 0 <
 0 | 0
 > 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sc2pir.res made by on Fri 19 May 95 8:48:10-PDT.

Query sequence being compared:	US-08-121-713B-2 (1-6)
Number of sequences searched:	75511
Number of scores above cutoff:	3756

Results of the initial comparison of US-08-121-713B-2 (1-6) with:
Data bank : PIR 43, all entries

100000-
-
-
N
U50000-
M
B
E
E
R
*
O
F10000-
-
-
S
E 5000-
Q
U
E
N
C
E

\$ 1000—
500—
100—
50—
10—
5—
0—

PARAMETERS									
Similarity matrix	Unitary	K-tuple							
Mismatch penalty	1	Joining penalty							2
Gap penalty	1.00	Window size							20
Gap size penalty	0.05								6
Cutoff score	0								
Randomization group	0								
Initial scores to save	45	Alignments to save							15
Optimized scores to save	0	Display context							10
SCORE 0	1	1	1	1	1	1	1	1	1
STDEV 0	1	2	3	3	3	4	4	5	6

SEARCH STATISTICS		
Scores:	Mean	Median
	1	3
		Standard Deviation
		1.02
Times:	CPU	Total Elapsed
	00:01:07.11	00:01:08.00
Number of residues:	22468934	
Number of sequences searched:	75511	
Number of scores above cutoff:	3756	

- Cut-off raised to 2.
- Cut-off raised to 3.
- Cut-off raised to 4.
- Cut-off raised to 5.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

15 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Opt. Score	Sig.	Frame
1. RNEGB	DNA-directed RNA polymerase (1082	6	6	4-90 0
2. JC1469	beta-adrenergic-receptor kina	688	6	6	4-90 0
3. A39336	beta-adrenergic-receptor kina	688	6	6	4-90 0
4. S21296	mating-type locus protein b6	410	6	6	4-90 0
5. S21295	mating-type locus protein b5	410	6	6	4-90 0
6. B32696	mating-type locus protein b2	410	6	6	4-90 0
7. A36671	mating-type locus protein b1	410	6	6	4-90 0
8. D32696	b3 protein - smut fungus (Ust	410	6	6	4-90 0
9. A32696	b1 protein - smut fungus (Ust	410	6	6	4-90 0
10. E36671	mating-type locus protein b6	410	6	6	4-90 0
11. D36671	mating-type locus protein b5	410	6	6	4-90 0
12. C36671	mating-type locus protein b3	410	6	6	4-90 0
13. A49915	GutQ homolog - Escherichia co	317	6	6	4-90 0
14. W6WL58	E6 protein - human papillomav	149	6	6	4-90 0
15. S28743	NADH dehydrogenase (ubiquinon	81	6	6	4-90 0

The list of other best scores is:

Sequence Name	Description	Length	Opt. Score	Sig.	Frame
16. S01763	gene X protein - Escherichia	12	4	4	2-94 0
17. A28514	dihydrolipoamide dehydrogenas	16	4	4	2-94 0
18. A49245	clustered asparagine-rich pro	17	4	4	2-94 0
19. F27393	spore coat protein, 34K - Bac	20	4	4	2-94 0
20. A28070	muti protein - Escherichia co	20	4	4	2-94 0
21. C26944	cytotoxic T-lymphocyte protei	20	4	4	2-94 0
22. B48186	ATP synthase beta 2 subunit -	23	4	4	2-94 0
23. B27393	spore coat protein, 59K - Bac	23	4	4	2-94 0
24. S24279	hypothetical protein 3 - porc	23	4	4	2-94 0
25. B44560	terephthalate 1,2-dioxygenase	25	4	4	2-94 0
26. S26754	ribosomal protein YML3, mitoc	28	4	4	2-94 0
27. S03947	hydrogen dehydrogenase (EC 1.	29	4	4	2-94 0
28. A22977	delta-endotoxin - Bacillus th	30	4	4	2-94 0
29. JS0339	hypothetical 4.4K protein - L	35	4	4	2-94 0
30. JU0109	S-carboxymethylcysteine synth	36	4	4	2-94 0
31. A38352	N5-(carboxyethyl)ornithine sy	37	4	4	2-94 0
32. A48368	N5,N10-methylenetetrahydrometh	38	4	4	2-94 0
33. A30010	cytochrome-c oxidase (EC 1.9.	42	4	4	2-94 0
34. A26929	sacQ protein - Bacillus liche	46	4	4	2-94 0
35. EGST	epidermal growth factor - rat	48	4	4	2-94 0
36. VDBPHK	kil protein - phage HK022	50	4	4	2-94 0
37. A45352	41K fiber protein - human ade	51	4	4	2-94 0
38. S08288	epidermal growth factor, low	51	4	4	2-94 0
39. E31439	ovomucoid, third domain - ele	51	4	4	2-94 0

- 40. INLDAT
- 41. A36141
- 42. ERAD65
- 43. JN0740
- 44. JN0739
- 45. S15871

lectin alpha chain - Tangier
Cop protein - Bacillus subtil
early E1A 6K protein - human
hypothetical 6.7K protein - p
hypothetical 6.4K protein - p
neurotoxin 5 - Indian cobra

1. US-08-121-713B-2 (1-6)

RNEGB DNA-directed RNA polymerase (EC 2.7.7.6) beta chain - Euglena gracilis chloroplast
#formal name chloroplast Euglena gracilis
30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 31-Mar-1993
S09210
#accessions
#authors
#journal
#title
The Euglena gracilis chloroplast rpoB gene. Novel gene organization and transcription of the RNA polymerase subunit operon.
#cross-references MUID:90245579
#accession : S09210
#molecule_type DNA
#residues 1-1082 #label YEP
##cross-references EMBL:X17191
##note
the authors translated the codon CGT for residue 132 as Gly

GENETICS
#gene
#genome
#introns
CLASSIFICATION
KEYWORDS
SUMMARY
SEQUENCE
rpoB
chloroplast
13/3; 89/2; 117/1; 154/3; 227/3; 241/1; 288/2; 732/3
#superfamily DNA-directed RNA polymerase beta chain
chloroplast; nucleotidyltransferase; transcription
#length 1082 #molecular-weight 124531 #checksum 6067

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNIX
| |
KNYSFFNLLGCLNEISQKNFKLIKRV
610 X 620 630

2. US-08-121-713B-2 (1-6)

JC1469 beta-adrenergic-receptor kinase (EC 2.7.1.126) 2 -
ENTRY JC1469 #type complete
TITLE beta-adrenergic-receptor kinase (EC 2.7.1.126) 2 - human
#formal name Homo sapiens #common name man
ORGANISM 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
DATE
ACCESSIONS JC1469

REFERENCE JC1469
#authors Parruti, G.; Ambrosini, G.; Salliese, M.; De Blasi, A.
#journal Biochem. Biophys. Res. Commun. (1993) 190:475-481
#title Molecular cloning, functional expression and mRNA analysis of human beta-adrenergic receptor kinase 2.

#accession JC1469
#molecule_type mRNA
#residues 1-688 #label PAR
#note the nucleotide sequence is not given in this paper
COMMENT This enzyme plays a role in the regulation of receptor-mediated immune functions.

KEYWORDS phosphotransferase
SUMMARY #length 688 #molecular-weight 79601 #checksum 6536
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX
| | |
70 X X 80

KIGFLFKDFCLNEINEAVPQVKFYE

3. US-08-121-713B-2 (1-6)
A39336 beta-adrenergic-receptor kinase (EC 2.7.1.126) - b

ENTRY A39336 #type complete
TITLE beta-adrenergic-receptor kinase (EC 2.7.1.126) - bovine
ORGANISM #formal name Bos primigenius taurus #common name cattle
DATE 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 23-Jun-1993

ACCESSIONS A39336
REFERENCE Benovic, J.L.; Onorato, J.J.; Arriza, J.L.; Stone, W.C.; Lohse, M.; Jenkins, N.A.; Gilbert, D.J.; Copeland, N.G.; Caron, M.G.; Lefkowitz, R.J.
#journal J. Biol. Chem. (1991) 266:14939-14946
#title Cloning, expression, and chromosomal localization of beta-adrenergic receptor kinase 2. A new member of the receptor kinase family.

#cross-references MUID:91332005
#accession A39336 preliminary
#status preliminary
#molecule_type mRNA
#residues 1-688 #label BEN
#cross-references GB:M73216

KEYWORDS phosphotransferase
SUMMARY #length 688 #molecular-weight 79803 #checksum 9052
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX
| | |

RIGFLFKDFCLNEINEAVPQVKFYE
70 X X 80

4. US-08-121-713B-2 (1-6)
S21296 mating-type locus protein b6 - smut fungus (Ustilago

ENTRY S21296 #type complete
TITLE mating-type locus protein b6 - smut fungus (Ustilago maydis)
ORGANISM #formal name Ustilago maydis #common name corn smut
DATE 19-Feb-1994; #sequence_revision 19-Feb-1994; #text_change 19-Feb-1994

ACCESSIONS S21296
REFERENCE S21295
#authors Kronstad, J.W.; Leong, S.A.
#submission submitted to the EMBL Data Library, July 1990
#accession S21296
#status preliminary
#residues 1-410 #label KRO
#cross-references EMBL:X53902

SUMMARY #length 410 #molecular-weight 46517 #checksum 7490
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX
| | |

PKLSKSFLECLNEIETEHEFLRDKVEH
10 X 20 30

5. US-08-121-713B-2 (1-6)
S21295 mating-type locus protein b5 - smut fungus (Ustilago

ENTRY S21295 #type complete
TITLE mating-type locus protein b5 - smut fungus (Ustilago maydis)
ORGANISM #formal name Ustilago maydis #common name corn smut
DATE 19-Feb-1994; #sequence_revision 19-Feb-1994; #text_change 19-Feb-1994

ACCESSIONS S21295
REFERENCE S21295
#authors Kronstad, J.W.; Leong, S.A.
#submission submitted to the EMBL Data Library, July 1990
#accession S21295
#status preliminary
#residues 1-410 #label KRO
#cross-references EMBL:X53901

SUMMARY #length 410 #molecular-weight 46304 #checksum 6080
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX
| | |

PNFSLTSEFLECLNEIEHEFLRDKEN
10 X 20 30

6. US-08-121-713B-2 (1-6)
B32696 mating-type locus protein b2 - smut fungus (Ustilago maydis)
B32696 #type complete
TITLE mating-type locus protein b2 - smut fungus (Ustilago maydis)
ORGANISM #formal name Ustilago maydis #common name corn smut
DATE 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 14-Jul-1994
ACCESSIONS B32696
REFERENCE A36671
#authors Kronstad, J.W.; Leong, S.A.
#journal Genes Dev. (1990) 4:1384-1395
#title The b mating-type locus of Ustilago maydis contains variable and constant regions.
#cross-references MUID:91032990
#accession B36671
#status preliminary
#molecule_type DNA
#residues 1-410 #label KRO
#note sequence not compared to nucleotide translation the nucleotide sequence is not given in this paper
#note
REFERENCE A32696
#authors Schulz, B.; Banuett, F.; Dahl, M.; Schlesinger, R.; Schaefer, W.; Martin, T.; Herskowitz, I.; Kahmann, R.
#journal Cell (1990) 60:295-306
#title The b alleles of U. maydis, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif.
#cross-references MUID:90124638
#accession B32696
#status preliminary
#molecule_type DNA
#residues 1-410 #label SCH
#cross-references GB:M58554
CLASSIFICATION #superfamily mating-type locus protein b1
SUMMARY #length 410 #molecular-weight 45957 #checksum 5041
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXX
| | |
PNFSLTSEFLECLNEIEHEFLRDKEN
10 X 20 30

7. US-08-121-713B-2 (1-6)
A36671 mating-type locus protein b1 - smut fungus (Ustilago maydis)
A36671 #type complete
TITLE mating-type locus protein b1 - smut fungus (Ustilago maydis)
ORGANISM #formal name Ustilago maydis #common name corn smut
DATE 12-Apr-1991 #sequence_revision 03-Apr-1992 #text_change

30-Sep-1993
ACCESSIONS A36671
REFERENCE A36671
#authors Kronstad, J.W.; Leong, S.A.
#journal Genes Dev. (1990) 4:1384-1395
#title The b mating-type locus of Ustilago maydis contains variable and constant regions.
#cross-references MUID:91032990
#accession A36671
#status preliminary
#molecule_type DNA
#residues 1-410 #label KRO
#note the authors translated the codon GAG for residue 100 as Asp

CLASSIFICATION #superfamily mating-type locus protein b1
SUMMARY #length 410 #molecular-weight 46197 #checksum 7183
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXX
| | |
PNFSLTSEFLECLNEIEHEFLRDKEN
10 X 20 30

8. US-08-121-713B-2 (1-6)
D32696 b3 protein - smut fungus (Ustilago maydis)
ENTRY D32696 #type complete
TITLE b3 protein - smut fungus (Ustilago maydis)
ORGANISM #formal name Ustilago maydis #common name corn smut
DATE 22-Jun-1990 #sequence_revision 28-Aug-1992 #text_change 30-Sep-1993
ACCESSIONS D32696
REFERENCE A32696
#authors Schulz, B.; Banuett, F.; Dahl, M.; Schlesinger, R.; Schaefer, W.; Martin, T.; Herskowitz, I.; Kahmann, R.
#journal Cell (1990) 60:295-306
#title The b alleles of U. maydis, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif.
#cross-references MUID:90124638
#accession D32696
#status preliminary
#molecule_type DNA
#residues 1-410 #label SCH
#cross-references GB:M58555
#note the authors translated the codon CCG for residue 194 as Arg

CLASSIFICATION #superfamily mating-type locus protein b1
SUMMARY #length 410 #molecular-weight 46323 #checksum 3467
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

```

X      X
CXNXIX
| | |
PKLSLSKFLECLNEIEHEFLRDKVEH
10  X 20 30

9. US-08-121-713B-2 (1-6)
   A32696    b1 protein - smut fungus (Ustilago maydis)
ENTRY      A32696    #type complete
TITLE      b1 protein - smut fungus (Ustilago maydis)
ORGANISM   #formal name Ustilago maydis #common name corn smut
DATE       22-Jun-1990 #sequence_revision 28-Aug-1992 #text_change
30-Sep-1993
ACCESSIONS A32696
REFERENCE  #authors  Schulz, B.; Banuett, F.; Dahl, M.; Schlesinger, R.; Schaefer,
W.; Martin, T.; Herskowitz, I.; Kahmann, R.
#journal   Cell (1990) 60:295-306
#title     The b alleles of U. maydis, whose combinations program
pathogenic development, code for polypeptides containing a
homeodomain-related motif.
#cross-references MUID:90124638
#accession  A32696
##status    preliminary
##molecule_type DNA
##residues  1-410 #label SCH
##cross-references GB:M58553
CLASSIFICATION #superfamily mating-type locus protein b1
SUMMARY       #length 410 #molecular-weight 46254 #checksum 7550
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X      X
CXNXIX
| | |
PNFSLISFLECLNEIEHEFLRDKGEN
10  X 20 30

10. US-08-121-713B-2 (1-6)
    E36671    mating-type locus protein b6 - smut fungus (Ustilago maydis)
ENTRY      E36671    #type complete
TITLE      mating-type locus protein b6 - smut fungus (Ustilago maydis)
ORGANISM   #formal name Ustilago maydis #common name corn smut
DATE       12-Apr-1991 #sequence_revision 12-Apr-1991 #text_change
30-Sep-1993
ACCESSIONS E36671
REFERENCE  #authors  Kronstad, J.W.; Leong, S.A.
#journal   Genes Dev. (1990) 4:1384-1395
#title     The b mating-type locus of Ustilago maydis contains variable
and constant regions.
#cross-references MUID:91032990
```

```

#accession  E36671    preliminary
##status    preliminary
##molecule_type DNA
##residues  1-410 #label KRO
##cross-references GB:X54071
CLASSIFICATION #superfamily mating-type locus protein b1
SUMMARY       #length 410 #molecular-weight 46317 #checksum 7490
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X      X
CXNXIX
| | |
PKLSLSKFLECLNEIEHEFLRDKVEH
10  X 20 30

11. US-08-121-713B-2 (1-6)
    D36671    mating-type locus protein b5 - smut fungus (Ustilago maydis)
ENTRY      D36671    #type complete
TITLE      mating-type locus protein b5 - smut fungus (Ustilago maydis)
ORGANISM   #formal name Ustilago maydis #common name corn smut
DATE       12-Apr-1991 #sequence_revision 12-Apr-1991 #text_change
30-Sep-1993
ACCESSIONS D36671
REFERENCE  #authors  Kronstad, J.W.; Leong, S.A.
#journal   Genes Dev. (1990) 4:1384-1395
#title     The b mating-type locus of Ustilago maydis contains variable
and constant regions.
#cross-references MUID:91032990
#accession  D36671
##status    preliminary
##molecule_type DNA
##residues  1-410 #label KRO
##cross-references EMBL:X54069
CLASSIFICATION #superfamily mating-type locus protein b1
SUMMARY       #length 410 #molecular-weight 46304 #checksum 6080
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X      X
CXNXIX
| | |
PNFSLISFLECLNEIEHEFLRDKGEN
10  X 20 30

12. US-08-121-713B-2 (1-6)
    C36671    mating-type locus protein b3 - smut fungus (Ustilago maydis)
ENTRY      C36671    #type complete
TITLE      mating-type locus protein b3 - smut fungus (Ustilago maydis)
```

ORGANISM #formal_name Ustilago maydis #common_name corn smut
DATE 12-Apr-1991 #sequence_revision 12-Apr-1991 #text_change
17-Feb-1994
ACCESSIONS C36671
REFERENCE A36671
#authors Kronstad, J.W.; Leong, S.A.
#journal Genes Dev. (1990) 4:1384-1395
#title The b mating-type locus of Ustilago maydis contains variable
and constant regions.
#cross-references MUID:91032990
#accession C36671
#status preliminary
#molecule_type DNA
#residues 1-410 #label KRO
#note sequence not compared to nucleotide translation
#note the nucleotide sequence is not given in this paper
CLASSIFICATION #superfamily mating-type locus protein b1
SUMMARY #length 410 #molecular-weight 46382 #checksum 3513

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXN XIX
| |
PKLSKFLKCLNIEHEFLDKVEH
10 X 20 30

13. US-08-121-713B-2 (1-6)
A49915 GutQ homolog - Escherichia coli

ENTRY #type complete
TITLE GutQ homolog - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change
18-Nov-1994
ACCESSIONS A49915
REFERENCE A49915
#authors Cieslewicz, M.J.; Steenbergen, S.M.; Vimr, E.R.
#journal J. Bacteriol. (1993) 175:8018-8023
#title Cloning, sequencing, expression, and complementation analysis
of the Escherichia coli K1 kps region 1 gene, kpsE, and
identification of an upstream open reading frame encoding a
protein with homology to GutQ.
#cross-references MUID:94075243
#contents K1
#accession A49915
#status preliminary
#molecule_type DNA
#residues 1-317 #label CIE
#note sequence inconsistent with nucleotide translation
#note sequence extracted from NCBI backbone
SUMMARY #length 317 #molecular-weight 34434 #checksum 5610
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90

Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXN XIX
| |
EAEKMKDKCLNVIGDQGGK
300 310 X

14. US-08-121-713B-2 (1-6)
W6WL58 E6 protein - human papillomavirus type 58

ENTRY #type complete
TITLE E6 protein - human papillomavirus type 58
ORGANISM #formal_name human papillomavirus type 58
#note host Homo sapiens (man)
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
24-Feb-1994

ACCESSIONS E36779
REFERENCE A36779
#authors Kirii, Y.; Iwamoto, S.; Matsukura, T.
#journal Virology (1991) 185:424-427
#title Human papillomavirus type 58 DNA sequence.
#cross-references MUID:92024102
#accession E36779
#molecule_type DNA
#residues 1-149 #label KIR
#cross-references GB:D90400
#note the translation of the nucleotide sequence is not given
in this paper

CLASSIFICATION #superfamily papillomavirus E6 protein
KEYWORDS early protein; zinc finger
SUMMARY #length 149 #molecular-weight 17794 #checksum 1553
SEQUENCE

Initial Score = 6 Optimized Score = 6 Significance = 4.90
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXN XIX
| |
GDTLEOTLKCLNEILIRCIICORPL
90 X 100 110

15. US-08-121-713B-2 (1-6)
S28743 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain

ENTRY #type complete
TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - blue
musssel mitochondrion (SGC4)
ORGANISM #formal_name mitochondrion Mytilus edulis #common_name blue
musssel
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993
ACCESSIONS S28743
REFERENCE S28743
#authors Hoffmann, R.J.; Boore, J.L.; Brown, W.M.

```

#journal      Genetics (1992) 131:397-412
#title       A novel mitochondrial genome organization for the blue
              mussel, Mytilus edulis.
#accession   S28743
#status      preliminary
#residues    1-81 ##label HOF
##cross-references EMBL:M83756
#length      81 ##molecular-weight 8952 ##checksum 8800
SUMMARY
SEQUENCE

```

```
Initial Score      = 6  Optimized Score = 6  Significance = 4.90
Residue Identity  = 50% Matches      = 3  Mismatches = 3
Gaps              = 0  Conservative Substitutions = 0
```

X X
 CXXIX
 | | |
 GVHAGLGQFSCLINEIVQKNYFTYQLV
 40 50 60
 > o <
 o | o IntelliGenetics
 > o <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq2spt.res made by on Fri 19 May 95 8:53:47-PDT.

Query sequence being compared: US-08-121-713B-2 (1-6)
 Number of sequences searched: 43470
 Number of scores above cutoff: 3751

Results of the initial comparison of US-08-121-713B-2 (1-6) with:
Data bank : Swiss-Prot 31, all entries

100000-
N
U50000-
M
B
E
E
R
O
F10000-
S
E5000-
Q
U
E
N
C
E
S1000-
500-

SCORE 0 1 2 3 4 5 5 6

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	6
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		

Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	1	3	0.91

```
Times:                                CPU                               Total Elapsed
00:00:43.98                          00:00:44.00
```

Number of residues:	15335248
Number of sequences searched:	43470
Number of scores above cutoff:	3751

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

10 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig.	Frame
1. RPOB_EUGGR	DNA-DIRECTED RNA POLYMERASE B	1082	6	6	5.47	0
2. ARK2_HUMAN	BETA-ADRENERGIC RECEPTOR KINA	688	6	6	5.47	0
3. ARK2_BOVIN	BETA-ADRENERGIC RECEPTOR KINA	688	6	6	5.47	0
4. ARK2_RAT	BETA-ADRENERGIC RECEPTOR KINA	688	6	6	5.47	0
5. B3_USTMA	MATING-TYPE LOCUS ALLELE B3 P	410	6	6	5.47	0
6. B6_USTMA	MATING-TYPE LOCUS ALLELE B6 P	410	6	6	5.47	0
7. B5_USTMA	MATING-TYPE LOCUS ALLELE B5 P	410	6	6	5.47	0
8. B2_USTMA	MATING-TYPE LOCUS ALLELE B2 P	410	6	6	5.47	0
9. B1_USTMA	MATING-TYPE LOCUS ALLELE B1 P	410	6	6	5.47	0
10. VE6_HPV58	E6 PROTEIN.	149	6	6	5.47	0

The list of other best scores is:

Sequence Name	Description	Length	Score	Init. Opt.	Sig.	Frame
11. COTB_BACSU	SPORE COAT PROTEIN B (FRAGMENT)	23	4	4	3.28	0
12. RM03_YEAST	MITOCHONDRIAL 60S RIBOSOMAL P	28	4	4	3.28	0
13. HOXY_NUCOP	NAD-REDUCING HYDROGENASE HOXS	29	4	4	3.28	0
14. ITR2_CUCOA	TRYPSIN INHIBITOR IIB (CSTI-I	32	4	4	3.28	0
15. SG2_RANRI	SECRETOGAGLIN II (FRAGMENT)	33	4	4	3.28	0
16. DEF7_RABIT	CORTICOSTATIN VI (CS-VI) (NEU	34	4	4	3.28	0
17. DHCO_LACLA	N5-(L-1-CARBOXYETHYL)-L-ORNIT	37	4	4	3.28	0
18. CU47_LACCO	BACTERIOICIN CURVATICIN FS47 (38	4	4	3.28	0
19. FUC3_RAT	FUCTININ 3 (FUCOSYLTRANSFERAS	39	4	4	3.28	0
20. PRE_BACLI	REGULATORY PROTEIN.	40	4	4	3.28	0
21. LPW_VTBPA	TRP OPERON LEADER PEPTIDE.	41	4	4	3.28	0
22. REP5_ECOLI	REP PROTEIN (E PROTEIN) (F4 P	43	4	4	3.28	0
23. CRIC_RHOSH	HYDROXYNEUROSPORENE DEHYDROE	44	4	4	3.28	0
24. DEQ3_BACLI	DEGRADATION ENZYME REGULATION	46	4	4	3.28	0
25. EGF_RAT	EPIDERMAL GROWTH FACTOR (EGF)	48	4	4	3.28	0
26. C555_BACAZ	CYTOCHROME C555 (FRAGMENTS)	50	4	4	3.28	0
27. ATP8_PODAN	ATP SYNTHASE PROTEIN 8 (EC 3.	50	4	4	3.28	0
28. IOVO_EUDELL	OVOMUCOID (FRAGMENT).	51	4	4	3.28	0
29. LECA_LATTI	LECTIN ALPHA CHAIN.	54	4	4	3.28	0
30. ATP8_NEUCR	ATP SYNTHASE PROTEIN 8 (EC 3.	54	4	4	3.28	0
31. EIA6_ADE05	EARLY EIA 6 KD PROTEIN.	55	4	4	3.28	0
32. EIA6_ADE02	EARLY EIA 6 KD PROTEIN.	55	4	4	3.28	0
33. COP6_STAAD	COP-6 PROTEIN.	55	4	4	3.28	0
34. PSBK_TOBAC	PHOTOSYSTEM II 4 KD REACTION	61	4	4	3.28	0
35. TXW5_NAJNA	WEAK NEUROTOXIN 5.	62	4	4	3.28	0
36. RS16_ORYSA	30S RIBOSOMAL PROTEIN S16.	62	4	4	3.28	0
37. CXH2_ASPSC	CYTOTOXIN HOMOLOG S3C2.	63	4	4	3.28	0
38. TXW9_NAJKA	WEAK TOXIN CM-9A.	64	4	4	3.28	0
39. NX51_BUNFA	SHORT NEUROTOXIN 1 (TOXIN V-I	64	4	4	3.28	0
40. TXW8_NAJNA	WEAK NEUROTOXIN 8.	65	4	4	3.28	0
41. TXW7_NAJNA	WEAK NEUROTOXIN 7.	65	4	4	3.28	0
42. TXW6_NAJNA	WEAK NEUROTOXIN 6.	65	4	4	3.28	0
43. TXW0_NAJNI	WEAK TOXIN CM-10.	65	4	4	3.28	0

44. RL33_MAIZE 50S RIBOSOMAL PROTEIN L33. 65 4 4 3.28 0
45. RL29_BACSU 50S RIBOSOMAL PROTEIN L29. 66 4 4 3.28 0

1. US-08-121-713B-2 (1-6)

RPOB_EUGGR DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6

ID	RPOB_EUGGR	STANDARD;	PRT;	1082 AA.
AC	P23579;			
DT	01-NOV-1991 (REL. 20, CREATED)			
DT	01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)			
DT	01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)			
DE	DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6).			
GN	RPOB.			
OS	EUGLENA GRACILIS.			
OG	CHLOROPLAST.			
OC	EUKARYOTA; PLANTIA; PHYCOPHYTA; EUGLENOPHYTA.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PRINGSHEIM Z;			
RM	90245579			
RA	YEPIZ-PLASCENCIA G.M., RADEBAUGH C.A., HALLICK R.B.;			
RL	NUCLEIC ACIDS RES. 18:1869-1878(1990).			
CC	-!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION			
CC	OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS			
CC	SUBSTRATES.			
CC	-!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +			
CC	RNA(N).			
CC	-!- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR			
CC	SUBUNITS: ALPHA, BETA, BETA', AND BETA''.			
DR	EMBL; X17191; CHEGRO.			
DR	PIR; S09210; RNEGB.			
KW	TRANSCRIPTION; DNA-DIRECTED RNA POLYMERASE; CHLOROPLAST.			
SQ	SEQUENCE 1082 AA; 124531 MW; 6191111 CN;			

Initial Score = 6 Optimized Score = 6 Significance = 5.47
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXX
| | |
KNYSFFNLLGCLNEISQKNFKLIKVV
610 X 620 630

2. US-08-121-713B-2 (1-6)

ARK2_HUMAN BETA-ADRENERGIC RECEPTOR KINASE 2 (EC 2.7.1.126) (

ID ARK2_HUMAN STANDARD; PRT; 688 AA.

AC P35626;

DT 01-JUN-1994 (REL. 29, CREATED)

DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)

DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)

DE BETA-ADRENERGIC RECEPTOR KINASE 2 (EC 2.7.1.126) (BETA-ARK-2).

GN ADREK2 OR BARK2.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RM PARUTTI G., AMBROSINI G., SALLESE M., DE BLASI A.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 190:475-481(1993).
CC -!- FUNCTION: SPECIFICALLY PHOSPHORYLATES THE AGONIST-OCCUPIED FORM
CC OF THE BETA-ADRENERGIC AND CLOSELY RELATED RECEPTORS.
CC -!- CATALYTIC ACTIVITY: ATP + [BETA-ADRENERGIC RECEPTOR] = ADP +
CC [BETA-ADRENERGIC RECEPTOR] PHOSPHATE.
CC -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
CC KINASES. STRONG, TO OTHER KINASES THAT PHOSPHORYLATES G-COUPLED
CC RECEPTORS.
CC EMBL; X69117; HSBADRK2.
DR HSSP; P05132; ICTP.
DR MIN; 109636; 11TH EDITION.
DR PROSITE; PS00107; PROTEIN KINASE ATP.
DR PROSITE; PS00108; PROTEIN KINASE ST.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING.
FT DOMAIN 1 196 N-TERMINAL.
FT DOMAIN 197 436 CATALYTIC.
FT DOMAIN 437 688 C-TERMINAL.
FT NP BIND 197 205 ATP (BY SIMILARITY).
FT BINDING 220 220 ATP (BY SIMILARITY).
FT ACT SITE 317 317 BY SIMILARITY.
SQ SEQUENCE 688 AA; 79677 MW; 2386685 CN;

Initial Score = 6 Optimized Score = 6 Significance = 5.47
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X X
CXNIX
| | |
KIGFLLEKDFCLNEINEAVPQVKFYE
70 X X 80

3. US-08-121-713B-2 (1-6)
ARK2_BOVIN BETA-ADRENERGIC RECEPTOR KINASE 2 (EC 2.7.1.126) (

ID ARK2_BOVIN STANDARD; PRT; 688 AA.
AC P26818;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE BETA-ADRENERGIC RECEPTOR KINASE 2 (EC 2.7.1.126) (BETA-ARK-2).
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; ARTIODACTYLA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RM 91332005
RA BENOVIC J.L., ONORATO J.J., ARRIZA J.L., STONE W.C., LOHSE M.,
RA JENKINS N.A., GILBERT D.J., COPELAND N.G., CARON M.G.,
RA LEFKOWITZ R.J.
RL J. BIOL. CHEM. 266:14939-14946(1991).
CC -!- FUNCTION: SPECIFICALLY PHOSPHORYLATES THE AGONIST-OCCUPIED FORM
CC OF THE BETA-ADRENERGIC AND CLOSELY RELATED RECEPTORS, PROBABLY
CC INDUCING A DESENSITIZATION OF THEM.

CC -!- CATALYTIC ACTIVITY: ATP + [BETA-ADRENERGIC RECEPTOR] = ADP +
CC [BETA-ADRENERGIC RECEPTOR] PHOSPHATE.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS; BRAIN, SPLEEN > HEART, LUNG >
CC KIDNEY.
CC -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
CC KINASES. STRONG, TO OTHER KINASES THAT PHOSPHORYLATES G-COUPLED
CC RECEPTORS.
CC EMBL; M73216; BTBARK.
DR FIR; A39336; A39336.
DR HSSP; P05132; ICTP.
DR PROSITE; PS00107; PROTEIN KINASE ATP.
DR PROSITE; PS00108; PROTEIN KINASE ST.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
KW MULTIGENE FAMILY.
FT DOMAIN 1 196 N-TERMINAL.
FT DOMAIN 197 436 CATALYTIC.
FT DOMAIN 437 688 C-TERMINAL.
FT NP BIND 197 205 ATP (BY SIMILARITY).
FT BINDING 220 220 ATP (BY SIMILARITY).
FT ACT SITE 317 317 BY SIMILARITY.
SQ SEQUENCE 688 AA; 79803 MW; 2384837 CN;

Initial Score = 6 Optimized Score = 6 Significance = 5.47
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X X
CXNIX
| | |
RIGFLLEKDFCLNEINEAVPQVKFYE
70 X X 80

4. US-08-121-713B-2 (1-6)
ARK2_RAT BETA-ADRENERGIC RECEPTOR KINASE 2 (EC 2.7.1.126) (

ID ARK2_RAT STANDARD; PRT; 688 AA.
AC P26819;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE BETA-ADRENERGIC RECEPTOR KINASE 2 (EC 2.7.1.126) (BETA-ARK-2).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RM 93019546
RA ARRIZA J.L., DAWSON T.M., SIMERLY R.B., MARTIN L.J., CARON M.G.,
RA SNYDER S.H., LEFKOWITZ R.J.;
RL J. NEUROSCI. 12:4045-4055(1992).
CC -!- FUNCTION: SPECIFICALLY PHOSPHORYLATES THE AGONIST-OCCUPIED FORM
CC OF THE BETA-ADRENERGIC AND CLOSELY RELATED RECEPTORS.
CC -!- CATALYTIC ACTIVITY: ATP + [BETA-ADRENERGIC RECEPTOR] = ADP +
CC [BETA-ADRENERGIC RECEPTOR] PHOSPHATE.
CC -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
CC KINASES. STRONG, TO OTHER KINASES THAT PHOSPHORYLATES G-COUPLED
CC RECEPTORS.
CC EMBL; M87855; RNBARK2.

DR HSP; P05132; ICTP.
DR PROSITE; P500107; PROTEIN KINASE ATP.
DR PROSITE; P500108; PROTEIN KINASE ST.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING.
FT DOMAIN 1 196 N-TERMINAL.
FT DOMAIN 197 436 CATALYTIC.
FT DOMAIN 437 688 C-TERMINAL.
FT NP BIND 197 205 ATP (BY SIMILARITY).
FT BINDING 220 220 ATP (BY SIMILARITY).
FT ACT SITE 317 317 BY SIMILARITY.
SQ SEQUENCE 688 AA; 79887 MW; 2345374 CN;

Initial Score = 6 Optimized Score = 5.47
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXX
| | |

KIGFLFKDFCINEIGEAIVPQVKEYE
70 X X 80

5. US-08-121-713B-2 (1-6)
B3_USTMA MATING-TYPE LOCUS ALLELE B3 PROTEIN.

ID B3_USTMA STANDARD; PRT; 410 AA.
AC P22017;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)
DE MATING-TYPE LOCUS ALLELE B3 PROTEIN.
OS USTILAGO MAYDIS (SMUT FUNGUS).
OC EUKARYOTA; FUNGI; BASIDIOMYCOTINA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RRK32;
RM 90124638
RA SCHULZ B., BANDETT F., DAHL M., SCHLESINGER R., SCHAEFER W.,
RA MARTIN T., HERSKOWITZ I., KARHANN R.;
RL CELL 60:295-306(1990).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 18604;
RM 91032990
RA KRONSTAD J.W., LEONG S.A.;
RL GENES DEV. 4:1384-1395(1990).
CC -!- FUNCTION: THE B LOCUS HAS AT LEAST 25 ALLELES, AND ANY COMBINATION
OF TWO DIFFERENT B ALLELES YIELDS A MULTIMERIC REGULATORY PROTEIN,
THAT ACTIVATES GENES RESPONSIBLE FOR THE PATHOGENICITY AND FOR THE
SEXUAL DEVELOPMENT OF THE FUNGUS WITHIN THE CORN PLANT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; M58555; UMB3.
DR PIR; D32696; D32696.
DR PIR; C36671; C36671.
KW DEVELOPMENTAL PROTEIN; DNA-BINDING; HOMEBOX; NUCLEAR PROTEIN.
FT DOMAIN 1 110 VARIABLE DOMAIN BETWEEN B ALLELES.
FT DOMAIN 111 410 HIGHLY CONSERVED BETWEEN B ALLELES.
FT DOMAIN 276 308 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 333 410 NOT ESSENTIAL FOR B3 FUNCTION.

FT DNA_BIND 134 179 HOMEBOX-LIKE.
SQ SEQUENCE 410 AA; 46323 MW; 894255 CN;
Initial Score = 6 Optimized Score = 5.47
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXX
| | |

PKLSLKFLECLNEIEHEFLRDKVEH
10 X 20 30

6. US-08-121-713B-2 (1-6)
B6_USTMA MATING-TYPE LOCUS ALLELE B6 PROTEIN.

ID B6_USTMA STANDARD; PRT; 410 AA.
AC P22020;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)
DE MATING-TYPE LOCUS ALLELE B6 PROTEIN.
OS USTILAGO MAYDIS (SMUT FUNGUS).
OC EUKARYOTA; FUNGI; BASIDIOMYCOTINA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 22505;
RM 91032990
RA KRONSTAD J.W., LEONG S.A.;
RL GENES DEV. 4:1384-1395(1990).
CC -!- FUNCTION: THE B LOCUS HAS AT LEAST 25 ALLELES, AND ANY COMBINATION
OF TWO DIFFERENT B ALLELES YIELDS A MULTIMERIC REGULATORY PROTEIN,
THAT ACTIVATES GENES RESPONSIBLE FOR THE PATHOGENICITY AND FOR THE
SEXUAL DEVELOPMENT OF THE FUNGUS WITHIN THE CORN PLANT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X54071; UMMTB6.
DR PIR; E36671; E36671.
DR PIR; S21296; S21296.
KW DEVELOPMENTAL PROTEIN; DNA-BINDING; HOMEBOX; NUCLEAR PROTEIN.
FT DOMAIN 1 110 VARIABLE DOMAIN BETWEEN B ALLELES.
FT DOMAIN 111 410 HIGHLY CONSERVED BETWEEN B ALLELES.
FT DOMAIN 276 308 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 333 410 NOT ESSENTIAL FOR B6 FUNCTION.
FT DNA_BIND 134 179 HOMEBOX-LIKE.
SQ SEQUENCE 410 AA; 46517 MW; 901339 CN;

Initial Score = 6 Optimized Score = 5.47
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXX
| | |

PKLSLKFLECLNEIEHEFLRDKVEH
10 X 20 30

7. US-08-121-713B-2 (1-6)
B5_USTMA MATING-TYPE LOCUS ALLELE B5 PROTEIN.

ID B5 USTMA STANDARD; PRT; 410 AA.
AC P22019;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DE MATING-TYPE LOCUS ALLELE B5 PROTEIN.
OS USTILAGO MAYDIS (SMUT FUNGUS).
OC EUKARYOTA; FUNGI; BASIDIOMYCOTINA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 22907;
RM 91032990
RA KRONSTAD J.W., LEONG S.A.;
RL GENES DEV. 4:1384-1395(1990).
CC -!- FUNCTION: THE B LOCUS HAS AT LEAST 25 ALLELES, AND ANY COMBINATION
OF TWO DIFFERENT B ALLELES YIELDS A MULTIMERIC REGULATORY PROTEIN,
THAT ACTIVATES GENES RESPONSIBLE FOR THE PATHOGENICITY AND FOR THE
SEXUAL DEVELOPMENT OF THE FUNGUS WITHIN THE CORN PLANT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X54069; UMMT85.
DR PIR; D36671; D36671.
DR PIR; S21295; S21295.
KW DEVELOPMENTAL PROTEIN; DNA-BINDING; HOMEBOX; NUCLEAR PROTEIN.
FT DOMAIN 1 110 VARIABLE DOMAIN BETWEEN B ALLELES.
FT DOMAIN 111 410 HIGHLY CONSERVED BETWEEN B ALLELES.
FT DOMAIN 276 308 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 333 410 NOT ESSENTIAL FOR B5 FUNCTION.
FT DNA BIND 134 179 HOMEBOX-LIKE.
SQ SEQUENCE 410 AA; 46304 MW; 903769 CN;

Initial Score = 6 Optimized Score = 6 Significance = 5.47
Residue Identity = 50% Matches = 3 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

X X
CXNXX
| |
PNFSLTSFLECLNEIEHEFLRKLEN
10 X 20 30

8. US-08-121-713B-2 (1-6)
B2 USTMA MATING-TYPE LOCUS ALLELE B2 PROTEIN.

ID B2 USTMA STANDARD; PRT; 410 AA.
AC P22016;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DE MATING-TYPE LOCUS ALLELE B2 PROTEIN.
OS USTILAGO MAYDIS (SMUT FUNGUS).
OC EUKARYOTA; FUNGI; BASIDIOMYCOTINA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=518;
RM 90124638
RA SCHULZ B., BANUETT F., DAHL M., SCHLESINGER R., SCHAEFER W.,
RA MARTIN T., HERSKOWITZ I., KAHMANN R.;
RL CELL 60:295-306(1990).

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=518;
RM 91032990
RA KRONSTAD J.W., LEONG S.A.;
RL GENES DEV. 4:1384-1395(1990).
CC -!- FUNCTION: THE B LOCUS HAS AT LEAST 25 ALLELES, AND ANY COMBINATION
OF TWO DIFFERENT B ALLELES YIELDS A MULTIMERIC REGULATORY PROTEIN,
THAT ACTIVATES GENES RESPONSIBLE FOR THE PATHOGENICITY AND FOR THE
SEXUAL DEVELOPMENT OF THE FUNGUS WITHIN THE CORN PLANT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; M38354; UMB2.
DR PIR; B32696; B32696.
DR PIR; B36671; B36671.
KW DEVELOPMENTAL PROTEIN; DNA-BINDING; HOMEBOX; NUCLEAR PROTEIN.
FT DOMAIN 1 110 VARIABLE DOMAIN BETWEEN B ALLELES.
FT DOMAIN 111 410 HIGHLY CONSERVED BETWEEN B ALLELES.
FT DOMAIN 276 308 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 333 410 NOT ESSENTIAL FOR B2 FUNCTION.
FT DNA BIND 134 179 HOMEBOX-LIKE.
SQ SEQUENCE 410 AA; 45957 MW; 899890 CN;

Initial Score = 6 Optimized Score = 6 Significance = 5.47
Residue Identity = 50% Matches = 3 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

X X
CXNXX
| |
PNFSLTSFLECLNEIEHEFLRKLEN
10 X 20 30

9. US-08-121-713B-2 (1-6)
B1 USTMA MATING-TYPE LOCUS ALLELE B1 PROTEIN.

ID B1 USTMA STANDARD; PRT; 410 AA.
AC P22015;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DE MATING-TYPE LOCUS ALLELE B1 PROTEIN.
OS USTILAGO MAYDIS (SMUT FUNGUS).
OC EUKARYOTA; FUNGI; BASIDIOMYCOTINA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=521;
RM 90124638
RA SCHULZ B., BANUETT F., DAHL M., SCHLESINGER R., SCHAEFER W.,
RA MARTIN T., HERSKOWITZ I., KAHMANN R.;
RL CELL 60:295-306(1990).

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=521;
RM 91032990
RA KRONSTAD J.W., LEONG S.A.;
RL GENES DEV. 4:1384-1395(1990).
CC -!- FUNCTION: THE B LOCUS HAS AT LEAST 25 ALLELES, AND ANY COMBINATION
OF TWO DIFFERENT B ALLELES YIELDS A MULTIMERIC REGULATORY PROTEIN,
THAT ACTIVATES GENES RESPONSIBLE FOR THE PATHOGENICITY AND FOR THE

CC	SEXUAL DEVELOPMENT OF THE FUNGUS WITHIN THE CORN PLANT.
CC	-I- SUBCELLULAR LOCATION: NUCLEAR.
DR	EMBL; M58553; UMB1.
DR	EMBL; X61305; UMBMTLRF.
DR	PIR; A32696; A32696.
DR	PIR; A36671; A36671.
KW	DEVELOPMENTAL PROTEIN; DNA-BINDING; HOMEBOX; NUCLEAR PROTEIN.
FT	DOMAIN 1 110 VARIABLE DOMAIN BETWEEN B ALLELES.
FT	DOMAIN 111 410 HIGHLY CONSERVED BETWEEN B ALLELES.
FT	DOMAIN 276 308 NUCLEAR LOCALIZATION SIGNAL
FT	(POTENTIAL).
FT	DOMAIN 333 410 NOT ESSENTIAL FOR B1 FUNCTION.
FT	DOMAIN BIND 134 179 HOMEBOX-LIKE.
FT	CONFLICT 168 169 HV -> QL (IN REF. 2).
FT	CONFLICT 218 218 Y -> S (IN REF. 2).
FT	CONFLICT 292 292 N -> K (IN REF. 2)..
SQ	SEQUENCE 410 AA; 46254 MW; 904704 CN;
Initial Score	= 6 Optimized Score = 6 Significance = 5.47
Residue Identity	= 50% Matches = 3 Mismatches = 3
Conservative Substitutions	= 0 Conservative Substitutions = 0

X X
CXNXIX
+ + +
PNFSLISFLECLNEIEHEFLRDKGEN
10 X 20 30

10.	DS-08-121-713B-2 (1-6)
	VE6 HPV58 E6 PROTEIN.
ID	VE6 HPV58 STANDARD; PRT; 149 AA.
AC	P26555;
DT	01-AUG-1992 (REL. 23, CREATED)
DT	01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT	01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE	E6 PROTEIN.
DE	E6.
GN	HUMAN PAPILLOMAVIRUS TYPE 58.
OS	VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAFOVAVIRIDAE; PAPILLOMAVIRUSES.
OC	[1]
RN	RP SEQUENCE FROM N.A.
RP	

92024102	
RM	
RA	KIRII Y., IWAMOTO S., MATSUKURA T.;
RL	VIROLOGY 185:424-427(1991).
CC	-!- FUNCTION: EXHIBIT A STRONG, BUT NON SPECIFIC AFFINITY FOR DOUBLE
CC	STRANDED DNA (IN VITRO).
CC	-!- SUBCELLULAR LOCATION: NUCLEAR MATRIX-ASSOCIATED.
DR	EMBL; D90400; PAPPFH58.
KW	PIR; E36779; W6WL58.
DR	EARLY PROTEIN; DNA-BINDING; NUCLEAR PROTEIN; ZINC-FINGER.
FT	ZN FING 30 66 POTENTIAL.
FT	ZN FING 103 139 POTENTIAL.
SQ	SEQUENCE 149 AA; 117794 MW; 106368 CN;

Initial Score	=	6	Optimized Score	=	6	Significance	=	5.47
Residue Identity	=	50%	Matches	=	3	Mismatches	=	3
Gaps	=	0	Conservative Substitutions	=	0		=	0

	X	X	
	CXNXIX		
GDTLEQTLKGCINEILRCIIICORPL			
90 X 100			
110			
11.	US-08-121-713B-2 (1-6)		
	COTB_BACSU SPORE COAT PROTEIN B (FRAGMENT) .		
ID	COTB_BACSU STANDARD; PRT; 23 AA.		
AC	P07789;		
DT	01-AUG-1988 (REL. 08, CREATED)		
DT	01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)		
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)		
DE	SPORE COAT PROTEIN B (FRAGMENT) .		
GN	COTB.		
OS	BACILLUS SUBTILIS.		
OC	PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RM	88011308		
RA	DONOVAN W., ZHENG L., SANDMAN K., LOSICK R.;		
RL	J. MOL. BIOL. 196:1-10(1987).		
DR	EMBL; X05679; BSCOTBG.		
DR	PIR; B27393; B27393.		
DR	SUBTILIST; BG10491; COTB.		
KW	SPORULATION.		
FT	NON TER 23 23		
SQ	SEQUENCE 23 AA; 2937 MW; 3201 CN;		
Initial Score =	4	Optimized Score =	4
Residue Identity =	33%	Matches =	2
Gaps =	0	Conservative Substitutions =	4
		Mismatches =	0

MSKRRMKYHSNNEISYYNFLHSM
10 X 20

	US-08-121-713B-2 (1-6)	RM03_YEAST	MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L3 (YML3)	(FRA
12.	US-08-121-713B-2 (1-6)	RM03_YEAST	MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L3 (YML3)	(FRA
ID	RM03_YEAST	STANDARD;	PRT;	28 AA.
AC	P36516;			
DT	01-JUN-1994	(REL. 29, CREATED)		
DT	01-JUN-1994	(REL. 29, LAST SEQUENCE UPDATE)		
DT	01-JUN-1994	(REL. 29, LAST ANNOTATION UPDATE)		
DE	MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L3 (YML3)			(FRAGMENT) .
DE	MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L3 (YML3)			
GN	MRP13.			
OS	SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).			
OC	EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.			
RN	[1]			
RP	SEQUENCE.			
RM	89078618			
RA	GRAACK H.-R., GROHMANN L., CHOLI T.;			
RL	FEBS LETT. 242:4-8 (1988).			
DR	PIR; S26754; S26754.			

KW RIBOSOMAL PROTEIN; MITOCHONDRION.
 FT NON TER 28
 SQ SEQUENCE 28 AA; 3387 MW; 4768 CN;
 Initial Score = 4 Optimized Score = 4 Significance = 3.28
 Residue Identity = 40% Matches = 2 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 CXNXIX
 I
 YKEYYQGLKSTVNEI
 20 X X

13. US-08-121-713B-2 (1-6)
 HOXY_NOCOP NAD-REDUCING HYDROGENASE HOXS DELTA SUBUNIT (EC 1.

ID HOXY NOCOP STANDARD; PRT; 29 AA.
 AC P22660;
 DT 01-AUG-1991 (REL. 19, CREATED)
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
 DT 01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)
 DE NAD-REDUCING HYDROGENASE HOXS DELTA SUBUNIT (EC 1.12.1.2) (FRAGMENT).
 GN HOXY.
 OS NOCARDIA OPACA.
 OG PLASMID.
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; NOCARDIOFORM.
 RN [1]
 RP SEQUENCE.
 RC STRAIN=1B;
 RM 89231684
 RA ZABOROSCH C., SCHNEIDER K., SCHLEGEL H.G., KRATZIN H.;
 RL EUR. J. BIOCHEM. 181:175-180(1989).
 CC -!- CATALYTIC ACTIVITY: H(2) + NAD(+) = H(+) + NADH.
 CC -!- COFACTOR: FMN, NICKEL, TWO 4FE-4S, A 3FE-4S, AND A 2FE-2S CLUSTER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SUBUNIT: Tetramer of an alpha and a gamma subunits (flavin-containing dimer), and a delta and a nickel-containing beta subunits (hydrogenase dimer).
 CC -!- SIMILARITY: TO OTHER [NIFE] OR [NIFESE] HYDROGENASES SMALL SUBUNIT.
 DR PIR: S03947; S03947.
 KW OXIDOREDUCTASE; IRON-SULFUR; 3FE-4S; PLASMID.
 FT NON TER 29
 SQ SEQUENCE 29 AA; 3159 MW; 4265 CN;

Initial Score = 4 Optimized Score = 4 Significance = 3.28
 Residue Identity = 33% Matches = 2 Mismatches = 4
 Gaps = 0 Conservative Substitutions = 0

X X
 CXNXIX
 I
 MKHSEKNEIASHELPTTPID
 X 10 20

14. US-08-121-713B-2 (1-6)

ITR2_CUCSA TRYPSIN INHIBITOR IIB (CSTI-IIB).

ID ITR2 CUCSA STANDARD; PRT; 32 AA.
 AC P10231;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-MAR-1989 (REL. 10, LAST ANNOTATION UPDATE)
 DE TRYPSIN INHIBITOR IIB (CSTI-IIB).
 OS CUCUMIS SATIVUS (CUCUMBER).
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 OC VIOLALES; CUCURBITACEAE.
 RN [1]
 RP SEQUENCE.
 RC TISSUE=SEED;
 RM 85149300
 RA WIECZOREK M., OTLEWSKI J., COOK J., PARKS K., LELOK J.,
 RA WILLIMOWSKA-PELC A., POLANOWSKI A., WILUSZ T., LASKOWSKI M. JR.;
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 126:646-652(1985).
 CC -!- SIMILARITY: BELONGS TO THE SQUASH FAMILY OF SERINE PROTEASE INHIBITORS.

DR HSSP: P01074; 1CTI.
 DR PROSITE: PS00286; SQUASH INHIBITOR.
 KW SERINE PROTEASE INHIBITOR.
 FT ACT SITE 5 6 REACTIVE BOND.
 FT DISULFID 3 20 BY SIMILARITY.
 FT DISULFID 10 22 BY SIMILARITY.
 FT DISULFID 16 29 BY SIMILARITY.
 SQ SEQUENCE 32 AA; 3535 MW; 5380 CN;

Initial Score = 4 Optimized Score = 4 Significance = 3.28
 Residue Identity = 33% Matches = 2 Mismatches = 4
 Gaps = 0 Conservative Substitutions = 0

X X
 CXNXIX
 I
 HSDCLDLCVCLDGYGCVGS
 20 X X 30

15. US-08-121-713B-2 (1-6)
 SG2_RANRI SECRETAGRANIN II (FRAGMENT).

ID SG2 RANRI STANDARD; PRT; 33 AA.
 AC P30345;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE SECRETAGRANIN II (FRAGMENT).
 OS RANA RIDIBUNDA (LAUGHING FROG) (MARSH FROG).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
 RN [1]
 RP SEQUENCE.
 RC TISSUE=BRAIN;
 RM 91285100
 RA VAUDRY H., CONLON J.M.;
 RL FEBS LETT. 284:31-33(1991).
 CC -!- FUNCTION: MAY BE IMPORTANT IN REGULATION OF NEUROSECRETION.
 CC -!- SIMILARITY: MEMBER OF THE CHROMOGRANIN/SECRETAGRANIN PROTEIN FAMILY.

DR PIR; S15867; S15867. 1
FT NON TER 1 33
FT PEPTIDE 1 33 BRAIN PEPTIDE.
FT NON TER 33 33
SQ SEQUENCE 33 AA; 3607 MW; 5190 CN;
Initial Score = 4 Optimized Score = 4 Significance = 3.28
Residue Identity = 40% Matches = 2 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CXNXIX

||
TNEIVEGQYTFQSLA
X X 10

maryh@stic

stdin

NeWSprinter20

Fri May 19 10:55:23 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5

11. R29939 Deduced from Lelystad Agent g 2396 7 2.65 0
12. R70373 Human fibronectin gene produc 2327 7 2.65 0
13. R15468 Human fibronectin. 2327 7 2.65 0
14. R81048 Sequence encoded by Rhinoviru 2150 7 2.65 0
15. R05127 Complete human rhinovirus 2- 2150 7 2.65 0
16. R60045 Sequence of viral proteins VP 2150 7 2.65 0
17. R95645 Rabbit skeletal muscle alpha-1 1873 7 2.65 0
18. R22461 Primary amino acid sequence o 1764 7 2.65 0
19. R22461 Masking protein high polymer 1712 7 2.65 0
20. R81184 Sequence encoded by the 2nd r 1594 7 2.65 0
21. R27640 Human calcium channel 27980/2 1493 7 2.65 0
22. R11749 Human alpha-2 macroglobulin b 1484 7 2.65 0
23. R11334 Recombinant human alpha-2 mac 1474 7 2.65 0
24. R37508 Human DNA polymerase alpha ca 1462 7 2.65 0
25. R24033 Soluble mannose receptor pept 1456 7 2.65 0
26. R38304 Sequence of a serrate protein 1404 7 2.65 0
27. R11061 Bovine Coronavirus E2 protein 1363 7 2.65 0
28. R14584 TGF beta 1 binding protein en 1355 7 2.65 0
29. R14481 LktA:lacZ fusion product. 1334 7 2.65 0
30. R30290 lktA:lacZ fusion protein fro 1334 7 2.65 0
31. R53787 Deep Vent DNA polymerase. 1312 7 2.65 0
32. R28337 SFV4 structural polyprotein. 1253 7 2.65 0
33. R45945 Glutamic acid receptor. 1239 7 2.65 0
34. R49832 Beat-galactosidase/hepatitis 1217 7 2.65 0
35. R26188 HVTa antigen. 1199 7 2.65 0
36. R32882 Cardiac adenylyl cyclase type 1184 7 2.65 0
37. R36728 Ubiquitin fusion protein, Ub- 1121 7 2.65 0
38. R38861 GC-C. 1075 7 2.65 0
39. R38096 Pyrococcus sp DNA polymerase. 1019 7 2.65 0
40. R29648 AmEPV Spheroidin protein GSR. 1003 7 2.65 0
41. R55576 AmEPV spheroidin. 1003 7 2.65 0
42. R30742 Human pemphigus vulgaris 130k 999 7 2.65 0
43. P50231 Sequence encoded by partial s 993 7 2.65 0
44. R25141 JAK2. 986 7 2.65 0
45. R42995 Glycosyltransferase. 985 7 2.65 0

1. US-08-121-713B-3 (1-7)
R47861 Alpha 2-Macroglobulin/LDL-receptor related protein

ID R47861 standard; protein; 4544 AA.
AC R47861;
DT 20-JUL-1994 (first entry)
DE Alpha 2-Macroglobulin/LDL-receptor related protein.
KW alpha-2 macroglobulin; Low Density Lipoprotein; LDL; receptor family;
KW LDL receptor related protein; LRP; small rhinovirus receptor; deriv;
OS Minor Rhinovirus; alpha2MR/LRP.
FH Homo sapiens.
FH Key Location/Qualifiers
FT Misc difference 211..260
FT /note= "50 residues not shown in SEQ.ID.No.4"
FT Misc difference 1990
FT /note= "Residue not shown in SEQ.ID.No.4"
FT Misc difference 3050
FT /note= "Residue not shown in SEQ.ID.No.4"
PN W09401553-A.
PD 20-JAN-1994.
PF 05-JUL-1993; E01728.
PR 08-JUL-1992; DE-222385.
PR 22-AUG-1992; DE-227892.

PR 19-FEB-1993; DE-305063.
PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.
PI Blaas D, Gruenberger M, Hofer F, Huettinger M, Kerjaschki D;
PI Kowalski H, Kuechler E, Machat H;
DR WPI; 94-035060/04.
PT New peptide derivs. of receptor for rhinovirus - of the small
PT receptor gp., and derived DNA, transformed cells and antibodies,
PT used e.g. to treat or prevent rhinovirus infection
PS Claim 5; Fig 2; 76pp; German.
CC Functional derivatives of members of the Minor Rhinovirus Receptor
CC group are claimed. The alpha-2 Macroglobulin/LDL-receptor related
CC protein of sequence R47861 (Herz et al. EMBO J. 7;4119-4127 (1988))
CC is a preferred parent receptor. The derive, which are preferably
CC soluble, extracellular forms of the native receptors, are useful
CC for treating and preventing viral (esp. rhinoviral) infections.
CC N.B. the SEQ.ID. listing includes a sequence (no.4) which differs
CC from the alpha2-MR/LRP sequence as indicated in the Features Table.
SQ Sequence 4544 AA,
SQ 249A; 277R; 249N; 393D; 0 B; 331C; 164Q; 235E; 0 Z; 376G; 129H;
SQ 196I; 305L; 173K; 72 M; 133F; 231P; 308S; 259T; 88 W; 131Y; 245V;
Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions

X X X
CGTXXX
|||
DCMDGSDDEACGTGVRTCPIDEFQCNN
3640 3650 X 3660

2. US-08-121-713B-3 (1-7)
R43662 DEN1-S275/90 (ECACC V92042111).

ID R43662 standard; Protein; 3396 AA.
AC R43662;
DT 16-MAY-1994 (first entry)
DE DEN1-S275/90 (ECACC V92042111).
KW Dengue haemorrhagic fever; DHF; dengue fever; DF; dengue shock syndrome;
KW DSS; DEN1 polypeptides.
OS Dengue Fever Virus Type 1 strain S275/90.
FH Key Location/Qualifiers
FT Protein 1..114
FT /label= C
FT Protein 15..114
FT /label= C'
FT Protein 115..205
FT /label= PreM
FT Protein 206..280
FT /label= M
FT Protein 281..774
FT /label= E
FT Protein 775..1128
FT /label= NS1
FT Protein 1129..1344
FT /label= NS2A
FT Protein 1345..1474
FT /label= NS2B
FT Protein 1475..2093

	/label= NS3	2094..2242	
FT	Protein		
FT	/label= NS4A		
FT	Protein	2243..2492	
FT	/label= NS4B		
FT	Protein	2493..3396	
FT	/label= NS5		
FT	W09322440-A.		
PD	11-NOV-1993.		
PF	28-APR-1993; CAO182.		
PR	29-APR-1992; GB-009243.		
PA	(UYSI-) UNIV SINGAPORE NAT.		
PI	Chan Y, Fu J, Tan B, Yap E, Tan Y;		
DR	WFI; 93-368799/46.		
DR	N-PSDB; Q51476.		
Pt	New Dengue virus type 1 strain - used to obtain prods. for		
Pt	detection, diagnosis, vaccines and treatment involving virus		
PS	Claim 9; Page 20-34; 55pp; English.		
CC	DEN1 virus, strain 5275/90 was isolated from the serum of a dengue		
CC	haemorrhagic fever (DHF) patient. RNA was isolated from the virus		
CC	and used to prepare cDNA encoding DEN1 polypeptides. Dengue Virus		
CC	Type I prods. can be used for detection, diagnosis, vaccines		
CC	(inactivated form) or treatment of DEN1 infections. The sequences		
CC	given in Q51477-86 are oligonucleotides used to prepare cDNA		
CC	fragments corresp. to Dengue virus proteins, by PCR.		
SQ	Sequence 3396 AA;		
SQ	Z36A; 194R; 124N; 148D; 0 B; 60 C; 109Q; 220E; 0 Z; 279G; 74 H;		
SQ	193I; 317L; 206K; 129M; 106F; 139P; 203S; 257T; 95 W; 74 Y; 233V;		
<hr/>			
Initial Score =	7 Optimized Score =	3 Significance = 2.65	
Residue Identity =	42% Matches =	7 Mismatches =	4
Gaps	=	0 Conservative Substitutions	= 0
<hr/>			
	X X X		
	CGTXXXX		
NSTHEMYVMSCGTGNISAVNMTRML			
2710 X 2720			2730
<hr/>			
3. US-08-121-713B-3 (1-7)			
R26052	APC gene product in familial adenomatous polyposis		
ID R26052 standard; Protein; 2843 AA.			
AC R26052;			
DT 28-JAN-1993 (first entry)			
DE APC gene product in familial adenomatous polyposis.			
KW neoplasm; cancer; oncogene; tumour; growth; detection; diagnosis;			
KW prognosis; treatment; sporadic colorectal carcinomas; ss.			
OS Homo sapiens			
PN W09213103-A.			
PD 06-AUG-1992.			
PF 16-JAN-1992; U00376.			
PR 16-JAN-1991; GB-000963.			
PR 08-AUG-1991; US-741940.			
PA (CANC-) CANCER INST.			
PA (ICIL) IMPERIAL CHEM IND PLC.			
PA (UYJO) UNIV.			
PA JOHNS HOPKINS.			
PA (UTAH) UNIV UTAR.			

Albertsen H, Anand R, Carlson ML, Groden JL, Hedge PJ;
Jostling, Kinzler KW, Markham A, Nakamura Y, Thilveris A;
Vogelstein B, White RL, Markham AF;
WPI: 92-284685/34.
N-PSDB; Q27234.
Detection of somatic and germ-line alterations of human APC gene
- used to diagnose, treat and study familial adenomatosis
polyposis and sporadic colorectal cancer
Disclosure; Page 47; 132pp; English.
This sequence is encoded by the APC (Adenomatous Polyposis Coli)
gene associated with tumorigenesis, found on chromosome 5q.
The sequence may be mutated by deletions/insertions, inversions, or
point mutations of the gene. The APC gene is expressed in most normal
tissues as well suggesting that APC is a tumour suppressor.
Sequence 2843 AA;
S0 176A; 163R; 158N; 149D; 0 B; 38 C; 154Q; 208E; 0 Z; 149G; 67 H;
S0 115I; 206L; 199K; 56 M; 47 F; 181P; 433S; 168T; 14 W; 50 Y; 112V;
Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X X
CGTXXXX
|||
SHSLTIYSNACGTLMLSNRPKQEA
680 690

4. US-08-121-713B-3 (1-7)
P60053 Sequence of von Willebrand factor (vWF).

ID P60053 standard; Protein; 2813 AA.
AC P60053;
DT 22-JUL-1991 (first entry)
DE Sequence of von Willebrand factor (vWF).
KW Vascular injury; platelet plug formation.
OS Homo sapiens.
PN EP-197592-A.
PD 15-OCT-1986.
PF 26-MAR-1986; 200518.
PR 01-APR-1985; NL-000961.
PA (VRIE-) STICHT VRIEND LANDS.
PA (FRIN-) STICHT FRINDER RANT.
PI Pannekoek H, Verwey CL, Diergaarde PJ, Hart MHL;
DR WPI; 86-273304/42.
DR N-PSDB; N60061.
PT Recombinant cDNA plasmid or phage - contg. C-DNA fragment which
codes for biological activity of human von Willebrand factor
Disclosure; Fig 3; 37pp; English.
CC vWF (glyco) protein having the AA sequence corresponding to the
nucleotide sequence of 2518-8667 or 295-2517 of N60061 is claimed.
CC Also claimed are new microorganisms, animal cell or human cell contg.
CC the recombinant cDNA plasmid or phage, e.g. strain E.coli DH1 contg.
CC the recombinant cDNA plasmid pSP8800vWF is deposited as C.B.S. No
CC 163.86.
SQ Sequence 2813 AA;
SQ 160A; 137R; 101N; 160D; 0 B; 217C; 142Q; 181E; 0 Z; 207G; 77 H;
SQ 97 I; 233I; 104K; 55 M; 93 F; 173P; 201S; 146T; 27 W; 81V; 221V;

* Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX

NCPKGQVYLCGTCPCNLTCRSLSPDE
660 670

5. US-08-121-713B-3 (1-7)
P60462 Sequence of human von Willebrand Factor (VWF) prec

ID P60462 standard; Protein; 2813 AA.

AC P60462;
DT 25-JUN-1991 (first entry)
DE Sequence of human von Willebrand Factor (VWF) precursor.
KW Chronic renal failure; therapy; factor VIII C.
OS Homo sapiens.
PN WO8606096-A.
PD 23-OCT-1986.
PF 10-APR-1986; U00760.
PR 11-APR-1985; US-722108.
PA (CHIL-) CHILDRENS MED CENT.
PI GINSBURG D, Orkin SH, Kaufman RJ;
DR WPI; 86-291663/44.
DR N-PSDB; N60404.

PT Pure von Willebrand Factor - produced using an expression vector
including a DNA sequence encoding functional VWF protein
PS Disclosure; Table 2, Pages 18-36A; 54pp; English.
CC cDNA clones pVWH33, pVWH5 and pVWE6 which span 9 kb pairs of DNA and
encompass the entire protein coding region of VWF, were selected to
construct full length cDNA (N60404). The pure VWF produced is useful
in the treatment of von Willebrand's disease (VWD) and the patients
with chronic renal failure whose abnormal bleeding times are
corrected by crude cryoprecipitate. Pure VWF can also be used to
carry, stabilise and improve the therapeutic efficacy of factor
VIII:C.

SQ Sequence 2813 AA;
SQ 154A; 143R; 98 N; 155D; 0 B; 234C; 133Q; 181E; 0 Z; 205G; 70 H;
SQ 95 I; 227L; 108K; 56 M; 89 F; 176P; 207S; 151T; 26 W; 79 Y; 226V;

Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX

NCPKGQVYLCGTCPCNLTCRSLSPDE
660 670

6. US-08-121-713B-3 (1-7)
RL3887 Inositol-3-phosphate binding peptide.

ID RL3887 standard; Protein; 2749 AA.
AC RL3887;

DT 27-NOV-1991 (first entry)
DE Inositol-3-phosphate binding peptide.
KW I-3-P; transformation; inositol.
PN J03183482-A.
PD 09-AUG-1991.
PF 14-DEC-1989; 324256.
PR 14-DEC-1989; JP-324256.
PA (KYOW) KYOWA HAKKO KOGYO KK.
DR WPI; 91-277584/38.
DR N-PSDB; Q13593.

PT New polypeptide having binding affinity to inositol-3-phosphate -
prepd. by culturing cell contg. recombinant plasmid comprising
DNA and vector DNA
PT Disclosure; Fig 2(1-3); 11pp; Japanese.
CC The sequence encoding this peptide may be included in a
plasmid/vector for transformation of a host cell and mass-prodn.
CC of the peptide.
SQ Sequence 2749 AA;
SQ 146A; 155R; 144N; 154D; 0 B; 60 C; 128Q; 229E; 0 Z; 142G; 71 H;
SQ 155I; 314L; 167K; 69 M; 123F; 99 P; 180S; 136T; 21 W; 71 Y; 185V;

Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX

IIFLMSFVNCGCTFTRGYRAMVLDFEF
2370 X 2380

7. US-08-121-713B-3 (1-7)
R60021 Fibrinogen-alpha.

ID R60021 standard; Protein; 2446 AA.

AC R60021;
DT 23-FEB-1995 (first entry)
DE Fibrinogen-alpha.
KW Tissue binding; tissue sealing; wound healing; vulnary;
KW tissue-binding domain; TBD; crosslinking domain; fibronectin;
KW fibrinogen-alpha; heparin-binding domain; collagen-binding domain;
KW cell-binding domain; hybrid protein.

OS Homo sapiens.

PH Key Location/Qualifiers
FT Domain 282...608
FT /label= collagen-binding domain
FT /note= "acts as tissue-binding domain of hybrid
protein"
FT Domain 1812..2171
FT /label= heparin-binding domain
FT /note= "acts as tissue-binding domain of hybrid
protein"

PN WO9416085-A.

PD 21-JUL-1994.

PF 30-DEC-1993; U12687.

PR 30-DEC-1992; US-998271.

PA (ZYMO) ZYMOGENETICS INC.

PI Irani MH;

DR WPI; 94-249231/30.

DR N-PSDB; Q70009.
PT New hybrid proteins for use in tissue sealing and wound healing -
PT comprising a tissue-binding domain from a protein covalently
PT linked to a crosslinking domain of another protein
PS Disclosure; Page 37-48; 87pp; English.
CC Hybrid proteins have a tissue-binding domain (TBD) from 1 protein
CC linked to a crosslinking domain from another protein. The TBD
CC comprises: aa 2-926, 928-1338 and especially 2-1336 of the sequence
CC given in R60019; the heparin-binding domain (aa 1812-2171 of R60021)
CC of fibronectin; the collagen-binding domain (aa 282-608 of R60021)
CC of fibronectin; or the cell-binding domain (aa 1357-1903 or 1532-
CC 1631 of R60020) of fibronectin. DNA encoding a fibronectin-
CC fibronogen hybrid is given in Q70007, and sequences for fibronectin
CC and fibrinogen-alpha in Q70008 and Q70009, respectively.
SQ Sequence 2446 AA;
SQ 99 A; 125R; 99 N; 123D; 0 B; 63 C; 131Q; 142E; 0 Z; 204G; 48 H;
SQ 119I; 136L; 78 K; 27 M; 54 F; 191P; 198S; 267T; 40 W; 103Y; 199V;
SQ 119I; 136L; 78 K; 27 M; 54 F; 191P; 198S; 267T; 40 W; 103Y; 199V;

Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||

SEGRDNRKWCSTQNYDADQKFGFCP
440 X 450 X 460

8. US-08-121-713B-3 (1-7)
R28582 HCV amino acid sequence contg. antibody reactive p

ID R28582 standard; protein; 2436 AA.
AC R28582;
DT 22-MAR-1993 (first entry)
DE HCV amino acid sequence contg. antibody reactive peptides.
KW Hepatitis C virus; detection; peptides.
OS Hepatitis C virus.

FH Key Location/Qualifiers
FT Peptide 828..838
FT /note= "claimed peptide reactive to HCV antibody"
FT Peptide 1236..1261
FT /note= "claimed peptide reactive to HCV antibody"
FT Peptide 1254..1275
FT /note= "claimed peptide reactive to HCV antibody"
FT Peptide 1270..1295
FT /note= "claimed peptide reactive to HCV antibody"
FT Peptide 1288..1312
FT /note= "claimed peptide reactive to HCV antibody"
FT Peptide 1305..1325
FT /note= "claimed peptide reactive to HCV antibody"
FT Peptide 1318..1340
FT /note= "claimed peptide reactive to HCV antibody"
FT Peptide 1236..1275
FT /note= "claimed peptide reactive to HCV antibody"
FT J04288097-A.

PN 13-OCT-1992.
PF 07-NOV-1991; 301705.
PR 07-NOV-1991; JP-301705.
PA (OLYU) OLYMPUS OPTICAL CO LTD.

DR WPI; 92-387721/47.
PT New peptides) - are reactive to an antibody against type C
PT hepatitis virus, used to detect virus in patients
PS Disclosure; Fig 1; 11pp; Japanese.
CC The peptides (claimed) can be used to detect hepatitis C virus (HCV)
CC antibody positive patients, post-transfusion hepatitis can be prevented
CC by screening a sample from a HCV antibody-positive patient with the
CC peptide.
SQ Sequence 2436 AA;
SQ 223A; 126R; 62 N; 105D; 0 B; 88 C; 75 Q; 105E; 0 Z; 189G; 52 H;
SQ 112I; 242L; 79 K; 45 M; 70 F; 174P; 168S; 185T; 53 W; 81 Y; 202V;
SQ 112I; 242L; 79 K; 45 M; 70 F; 174P; 168S; 185T; 53 W; 81 Y; 202V;

Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||

GFLTNSRGCGTTRCRASGVLTTCG
2240 X 2250 2260

9. US-08-121-713B-3 (1-7)
R25135 HCV polypeptide 1.

ID R25135 standard; Protein; 2435 AA.
AC R25135;
DT 23-DEC-1992 (first entry)
DE HCV polypeptide 1.
KW Hepatitis C virus; blood transfusion.
OS Synthetic.
PN J04159298-A.
PD 02-JUN-1992.
PF 19-OCT-1990; 282431.
PR 19-OCT-1990; JP-282431.
PA (OLYU) OLYMPUS OPTICAL CO LTD.
DR WPI; 92-231947/28.

PT New peptides acting as antigenic analogues of human hepatitis C
PT virus - useful for detecting HCV antibody positive patients
PS Claim 1; Page 1; 14pp; Japanese.
CC The sequences given in R25135-36 are peptides from the hepatitis C
CC virus (HCV) which are recognised by the peptide sequences given in
CC R25130-24. These antigenic peptides can be used on their own or as
CC a mixture of two different peptides. Using these peptides, HCV antibody
CC positive patients can be detected and hepatitis caused by blood
CC transfusion can be prevented.

SQ Sequence 2435 AA;
SQ 223A; 126R; 62 N; 105D; 0 B; 88 C; 76 Q; 106E; 0 Z; 188G; 52 H;
SQ 112I; 239L; 80 K; 45 M; 70 F; 174P; 168S; 185T; 53 W; 81 Y; 202V;
SQ 112I; 239L; 80 K; 45 M; 70 F; 174P; 168S; 185T; 53 W; 81 Y; 202V;

Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||

GFLTNSRGCGTTRCRASGVLTTCG
2240 X 2250 2260

10. US-08-121-713B-3 (1-7)
R24306 Translation of ORF 2 contg. E.faecium protein Vans

ID R24306 standard; Protein; 2408 AA.
AC R24306;
DT 20-NOV-1992 (first entry)
DE Translation of ORF 2 contg. E.faecium protein Vans.
KW Glycopeptide antibiotic; vancomycin; teicoplanin; resistant;
KW D-Ala-D-Ala ligase; peptidoglycan precursor; transposon;
KW inverted repeats; vanR; vanS; vanH; vanA; vanX;
KW open reading frame.
OS Enterococcus faecium BM4147.
FH Key Location/Qualifiers
FT Protein 727..1115
FT /label= Vans
FT /note= "see R24296"
FN W09207942-A.
PD 14-MAY-1992.
PF 29-OCT-1991; F00855.
PR 31-OCT-1990; FR-013579.
PA (INSP) INST PASTEUR.
PI Arthur M. Courvalin P, Dutka-malen S, Molinas C;
DR WPI; 92-183677/22.
DR N-PSDB; Q25183.
PT Polypeptides involved in expression of glycopeptide antibiotic resistance - useful in diagnosing presence of Gram-positive enterococcal strains e.g. Enterococcus faecium and E gallinarum
PT Disclosure; Fig 5; 163pp; French.
CC A 7.3kb fragment of E.faecium DNA containing the five genes vanH, vanA, vanX, vanR and vanS involved in antibiotic resistance was translated in each of the three possible open reading frames.
CC Within ORF 2 there is the Vans gene. The "X"s in the sequence indicate the position of nonsense codons.
CC See also Q25178-Q25182.
SQ Sequence 2408 AA;
SQ 120A; 200R; 100N; 70 D; 0 B; 78 C; 82 Q; 66 E; 0 Z; 118G; 52 H;
SQ 182I; 212L; 171K; 50 M; 97 F; 82 P; 198S; 120T; 28 W; 115Y; 110V;
SQ 157 Others;

Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 57% Matches = 4 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
||||
QIRAFQSMYOCGTIXIRDFRLYSACEE
1230 1240 1250

11. US-08-121-713B-3 (1-7)
R29939 Deduced from Lelystad Agent genome ORF 1A.

ID R29939 standard; Protein; 2396 AA.
AC R29939;
DT 28-APR-1993 (first entry)
DE Deduced from Lelystad Agent genome ORF 1A.
KW Mystery Swine Disease; MSD; Arteriviridae; coronavirus; RNA genome;

KW swine infertility and respiratory syndrome; CDI-NL-2.91;
OS Abortus Blauw; viral replicase; polymerase; open reading frame.
FH Lelystad Agent.

FT Key Location/Qualifiers
FT Modified site 257..259
FT /label= N-glycosylation_site
FT /note= "putative"
FT Modified site 842..844
FT /label= N-glycosylation_site
FT /note= "putative"
FT Modified site 2081..2083
FT /label= N-glycosylation_site
FT /note= "putative"
FN W09221375-A.
PD 10-DEC-1992.
PF 05-JUN-1992; N10096.
PR 06-JUN-1991; EP-201398.
PR 18-MAR-1992; EP-200781.
PA (DIER-) STICHTING CENT DIERGENEESKUNDIG INST.
PI Meulenbergh JUM, Moormann RJM, Pol JMA, Terpstra C;
PI Wensvoort G;
DR WPI; 92-433379/52.
DR N-PSDB; Q32002.
PT Comph. contg. source agent of unknown Swine disease - in live, attenuated, dead or recombinant form, vaccine compans. and diagnostic kits, also causing agent specific nucleotide sequence
PS Claim 11; Fig 1; 77pp; English.
CC The genome of the Lelystad Agent consists of a genomic RNA molecule (19.5 to 15.5kb) which replicates via a 3' nested set of subgenomic RNAs. The nucleotide sequence was determined from overlapping cDNA clones. The complete nucleotide sequence and the proteins deduced from the 8 open reading frames (see R29939-R29946) are suitable for vaccine development. Diagnosis of MSD will also be possible. ORF 1A and ORF 1B are predicted to encode the viral replicase or polymerase, ORFs 2-6 are predicted to encode structural viral membrane (envelope) associated proteins and ORF 7 is predicted to encode the structural viral nucleocapsid protein. The products deduced from ORFs 6 and 7 show significant similarity with VpX and Vp1, respectively, of Lactate Dehydrogenase-Elevating Virus (LDV). The Lelystad Agent and LDV have been classified with Equine Arteritis Virus in the new virus family, the Arteriviridae which is part of the superfamily of coronaviruses.
SQ Sequence 2396 AA;
SQ 209A; 128R; 66 N; 115D; 0 B; 72 C; 85 Q; 105E; 0 Z; 183G; 54 H;
SQ 82 I; 256L; 89 K; 41 M; 102F; 173P; 199S; 139T; 46 W; 46 Y; 206V;

Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
||||
TGFILCHVEHCGTSGDSSPLDLSDA
810 820 X 830

12. US-08-121-713B-3 (1-7)
P70373 Human fibronectin gene product.

ID P70373 standard; protein; 2327 AA.

AC P70373;
DT 11-MAR-1991 (first entry)
DE Human fibronectin gene product.
KW FN; collagen; fibrin; heparin.
OS Homo sapiens.
PN EP-207751-A.
PD 07-JAN-1987.
PF 27-JUN-1986; 304998.
PR 28-JUN-1985; GB-016421.
PA (DELTA-) DELTA BIOTECHN.
PI Baralle FE;
DR WFI; 87-001441/01.
DR N-PSDB; N70596.
PT New fibronectin polypeptide sequence with affinity for collagen
PT etc. - useful for targeting therapeutic substances on natural
PT fibrin, for use in affinity purificn. of polypeptide(s) etc.
PS Claim 11; Fig 3A; 32pp; English.
CC The product may be expressed from a transformed micro-organism,
CC esp. E.coli.
CC FN binds to fibrin, heparin and Staphylococcus aureus, and may be
CC used to target a therapeutic agent onto natural fibrin eg. a blood
CC clot. It may also be used in affinity purification of a polypeptide,
CC conjugated to the collagen binding site of FN and immobilised on a
CC collagen surface.
SQ Sequence 2327 AA;
SQ 93 A; 122R; 97 N; 117D; 0 B; 62 C; 128Q; 137E; 0 Z; 190G; 48 H;
SQ 108I; 122L; 77 K; 26 M; 52 F; 184P; 185S; 253T; 39 W; 98 Y; 189V;
Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
CGTXXX
|||
SEGRDRNMKWCQTQNYDADQKFGFCP
410 420 X 430
13. US-08-121-713B-3 (1-7)
R15468 Human fibronectin.
ID R15468 standard; Protein; 2327 AA.
AC R15468;
DT 12-MAR-1992 (first entry)
DE Human fibronectin.
KW Fibrin-imag; atherosclerosis; thrombus inhibitor.
OS Homo sapiens.
PN WO9117765-A.
PD 28-NOV-1991.
PF 21-MAY-1991; U03584.
PR 21-MAY-1990; US-526397.
PA (BIOT-) BIO-TECHN GEN CORP.
PI Vogel T, Levanon A, Werber M, Guy R, Panet A, Hartman J;
PI Shaked H.
DR WFI; 91-369004/50.
DR N-PSDB; Q15214.
PT New fibrin binding domain polypeptide(s) - useful in imaging
PT fibrin-contg. substances, to inhibit thrombus formation and treat
PT wounds

PS Disclosure; Fig 2; 191pp; English.
CC The amino acid sequence is that of human fibronectin, this can be used
CC to derive polypeptides which are identical to part of the fibrin-
CC binding domain (FBD) of fibronectin. These polypeptides can be used
CC to inhibit thrombus formation; or (coupled to a thrombolytic agent)
CC to induce thrombolysis, or to treat wounds, e.g. in skin, eyes or
CC tendons (in conjunction with a polypeptide which includes a part of
CC the cell-binding domain (CBD) of fibronectin). These polypeptides are
CC easier to prepare than the full 31kD polypeptide. It can also be
CC used to image fibrin-contg. materials, esp. a thrombus or athero-
CC sclerotic plaque, pref. using a gamma counter.
SQ Sequence 2327 AA;
SQ 93 A; 122R; 97 N; 117D; 0 B; 62 C; 127Q; 137E; 0 Z; 190G; 49 H;
SQ 108I; 122L; 77 K; 26 M; 52 F; 184P; 186S; 252T; 39 W; 98 Y; 189V;
Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
X X
CGTXXX
|||
SEGRDRNMKWCQTQNYDADQKFGFCP
410 420 X 430
14. US-08-121-713B-3 (1-7)
P81048 Sequence encoded by Rhinovirus 2 (HRV2) genome cDN
ID P81048 standard; protein; 2150 AA.
AC P81048;
DT 22-MAR-1991 (first entry)
DE Sequence encoded by Rhinovirus 2 (HRV2) genome cDNA
KW Passive immunity; diagnosis; therapy; ss.
OS Rhinovirus.
FH Key Location/Qualifiers
FT Protein 1..69
FT /label=VP4 70..330
FT Protein
FT /label=VP2 331..567
FT Protein
FT /label=VP3 568..856
FT Protein
FT /label=VP1 857..882
FT Protein
FT /label=?
FT /note="protein named '??'"
FT Protein 883..992
FT /label=P2-A 993..1087
FT Protein
FT /label=P2-B 1088..1409
FT Protein
FT /label=P2-C 1410..1486
FT Protein
FT /label=P3-A 1487..1507
FT Protein
FT /label=VPg 1508..1690
FT Protein
FT /label=PROTEASE 1691..2150
FT Protein

FT /label=POLYMERASE
FN EP-261403-A.
PF 30-MAR-1988.
PD 20-AUG-1987; 112104.
PR 17-JAN-1987; DE-701301.
PA (BOEH) Boehringer Ingelheim.
PI Duechler M, Skern I, Sommergruber W, Neubauer C, Grundler P, Blaas D, Kuchler E, Fraese L, Zorn M.
DR WPI; 88-085735/13.
DR N-PSDB; N81393.
PT New DNA corresponding to viral RNA of rhino-virus HRV89 - useful for prodn. of polypeptide(s) for stimulating immune system against HRV 89.
PS Example; Fig 14; 66pp; German..
CC DNA molecules corresponding to all or part of the RNA of rhinovirus strain HRV89 (Fig 4, N81390) is claimed, esp. the portion encoding the viral proteins VP1-VP4, P2A-P2C, P3A-P3C. Also claimed are the polypeptides encoded by any of these DNA molecules. The polypeptides are used for stimulating a protective immune response and for blocking cellular receptors. Ab are useful for assay and for confirn. of the corresp. antigen, and can also be used for the therapeutic and diagnostic applications.
SQ Sequence 2150 AA;
SQ 119A; 86 R; 120N; 124D; 0 B; 48 C; 71 Q; 102E; 0 Z; 124G; 60 H; 181I; 163L; 133K; 49 M; 89 F; 120P; 162S; 146T; 29 W; 93 Y; 131V;
Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||
TCSLRFSFMFCGTANTTVKLLIAYTPP
440 450 X 460

15. US-08-121-713B-3 (1-7)
R05127 Complete human rhinovirus 2.

ID R05127 standard; protein; 2150 AA.

AC R05127;
DT 17-JUL-1990 (first entry)
DE Complete human rhinovirus 2.

KW Human rhinovirus 2.

FH Key Location/Qualifiers

FT Peptide	1..69
FT /label=VP4	70..330
FT Peptide	331..567
FT /label=VP2	568..852
FT Peptide	853..967
FT /label=VP1	993..1087
FT /label=P2-A	1088..1486
FT Peptide	
FT /label=P2-B	
FT Peptide	
FT /label=P2-C	

FT Peptide 1487..1507
FT /label=VPg
FT Protein 1508..1690
FT /label=Protease
FT Protein 1691..2150
FT /label=Polymerase
FN WO9001061-A.
PD 08-FEB-1990.
PF 22-JUL-1989; 0000861.
PR 24-JUN-1989; DE-920753.
PR 24-JUN-1989; DE-825189.
PA (BOEH) Boehringer Ingelheim.
PI Duechler M, Skern I, Blaas D, Berger B, Sommergruber W, Keuchler E.
DR WPI; 90-067175/09.
DR N-PSDB; Q03418.
PT New plasmid containing complete cDNA of human rhinovirus 2 under control of RNA polymerase promoter, transcribable in vitro to infectious RNA. Disclosure; Fig 3A-3E; 33pp; German.
CC The HRV2 cDNA transcribed to infectious RNA allows study of viral behaviour. Effects of deletions, insertions or exchanges of amino acids can be examined directly and quantitatively. Viral polypeptides expressed by these plasmids can be used therapeutically, eg. to stimulate the immune system.
SQ Sequence 2150 AA;
SQ 119 A; 86 R; 120 N; 124 D; 0 B; 48 C; 71 Q; 102 E; 0 Z; 124 G; 60 H; 181 I; 163 L; 133 K; 49 M; 89 F; 120 P; 162 S; 146 T; 29 W; 93 Y; 131 V;
Initial Score = 7 Optimized Score = 7 Significance = 2.65
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||
TCSLRFSFMFCGTANTTVKLLIAYTPP
440 450 X 460
> O <
O | O IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq3pir.res made by on Fri 19 May 95 8:52:13-PDT.

Query sequence being compared:US-08-121-713B-3 (1-7)
Number of sequences searched: 75311
Number of scores above cutoff: 4999

Results of the initial comparison of US-08-121-713B-3 (1-7) with:
Data bank : PIR 43, all entries

100000-	-
N	-
U50000-	-
M	-
B	-
E	-

*

A log-linear plot showing the distribution of scores for various categories. The y-axis is labeled 'SCORE' and ranges from 0 to 1000 on a logarithmic scale. The x-axis is labeled 'STDEV' and ranges from 0 to 7. The plot shows a series of points connected by lines, with a peak around STDEV 2.5. Asterisks are placed above the points at STDEV 2.5 and STDEV 6.5.

Category	STDEV	SCORE
R	0.0	10
O	0.5	15
F10000	1.0	20
S	1.5	30
E 5000	2.0	40
Q	2.5	50
U	3.0	60
E	3.5	70
N	4.0	80
C	4.5	90
E	5.0	100
S 1000	5.5	110
	6.0	120
	6.5	130
	7.0	140

SEARCH STATISTICS

Scores:	Mean	3	Median	6	Standard Deviation	2.18
Times:	CPU	00:01:08.06			Total Elapsed	00:01:09.00

Number of residues:	22468834
Number of sequences searched:	75511
Number of scores above cutoff:	4999

Cut-off raised to 5.
Cut-off raised to 6.

The scores below are sorted by initial score. Significance is calculated based on initial score.

1473 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
1. IUFTFM	cadherin-related tumor suppressor	5147	7	7	1.84
2. S02392	LDL receptor-related protein	4544	7	7	1.84
3. S25021	probable polyketide synthase	4427	7	7	1.84
4. S28600	hypothetical protein a - huma	4085	7	7	1.84
5. A40701	taxacin-X precursor - human	3566	7	7	1.84
6. A46112	genome polyprotein - rice tun	3473	7	7	1.84
7. S27927	genome polyprotein - rice tungro bac	3473	7	7	1.84
8. A42551	genome polyprotein - dengue v	3396	7	7	1.84
9. GNWVD3	genome polyprotein - dengue v	3390	7	7	1.84
10. S37536	macrogolgin - human	3259	7	7	1.84
11. S27852	hypothetical protein DGF-1 -	3229	7	7	1.84
12. A48450	putative cell-surface protein	3029	7	7	1.84
13. A44062	genome polyprotein - pepper m	3068	7	7	1.84
14. JQ1917	polyprotein - parsnip yellow	3027	7	7	1.84
15. S39658	polyposis coli protein DP2.5	2844	7	7	1.84
16. RBHUAP	adenomatous polyposis coli pr	2843	7	7	1.84
17. VWHU	von Willebrand factor precurs	2813	7	7	1.84
18. ACSMIT	inositol-trisphosphate recept	2749	7	7	1.84
19. A36579	inositol 1,4,5-trisphosphate r	2749	7	7	1.84
20. B36579	inositol 1,4,5-trisphosphate r	2734	7	7	1.84
21. A23475	G surface protein - Paramediu	2718	7	7	1.84
22. A34203	DNA-binding protein PRDII-BF1	2717	7	7	1.84
23. S28261	kinesin-related protein CENP-	2663	7	7	1.84
24. VF1HB2	RNA-directed RNA polymerase (2652	7	7	1.84
25. A30788	mannose 6-phosphate receptor	2499	7	7	1.84
26. A28372	mannose 6-phosphate receptor	2491	7	7	1.84
27. A43908	fibronectin - African clawed	2481	7	7	1.84
28. S14428	fibronectin precursor - rat	2477	7	7	1.84
29. RNQ2L	DNA-directed RNA polymerase (2452	7	7	1.84
30. B36861	RNA-directed RNA polymerase	2396	7	7	1.84
31. FNHU	fibronectin precursor - human	2386	7	7	1.84
32. S48405	hypothetical protein - yeast	2376	7	7	1.84
33. B47447	Ca2+ channel alpha 1 subunit	2326	7	7	1.84
34. FNBO	fibronectin - bovine	2259	7	7	1.84
35. S29236	calcium channel protein B11-1	2259	7	7	1.84

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	6
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

```
36. S06188 genome polyprotein 1 - grapev 2252 7 7 1.84 0
37. A47447 Ca2+ channel alpha 1 subunit 2223 7 7 1.84 0
38. A37490 calcium channel alpha 1 subun 2222 7 7 1.84 0
39. A41098 calcium channel protein alpha 2212 7 7 1.84 0
40. J01977 glutamate synthase (NADH) (EC 2194 7 7 1.84 0
41. S29237 calcium channel protein BII-2 2178 7 7 1.84 0
42. GNNY1B genome polyprotein - human rh 2157 7 7 1.84 0
43. GNNYH2 genome polyprotein - human rh 2150 7 7 1.84 0
44. A35672 crumbs protein - fruit fly (D 2139 7 7 1.84 0
45. AGCH agrin precursor - chicken 1955 7 7 1.84 0

1. US-08-121-713B-3 (1-7)
  IUFFTM cadherin-related tumor suppressor precursor - fruit fly
  ENTRY IUFFTM #type complete
  TITLE cadherin-related tumor suppressor precursor - fruit fly
  ORGANISM (Drosophila melanogaster)
  DATE #formal_name Drosophila melanogaster
  30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
  08-Dec-1994
  A41087; B41087
  ACCESSIONS A41087; B41087
  REFERENCE Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biessmann, H.;
  Bryant, P.J.; Goodman, C.S.
  #journal Cell (1991) 67:853-868
  #title The fat tumor suppressor gene in Drosophila encodes a novel
  member of the cadherin gene superfamily.
  #cross-references MUID:92069752
  #accession A41087
  ##molecule_type mRNA
  ##residues 143-485;1279-5147 ##label MAH
  ##cross-references GB:M80537
  #accession B41087
  ##molecule_type DNA
  ##residues 1-142;487-1278 ##label MA2
  ##cross-references GB:M80537
  ##note 1229-Gly and 1233-Ser were also found

GENETICS fat
#gene #superfamily cadherin-related tumor suppressor; cadherin
#CLASSIFICATION repeat homology; EGF homology
#KEYWORDS calcium binding; cell adhesion; duplication; transmembrane
protein
#domain signal sequence #status predicted #label SIG\
#product cadherin-related tumor suppressor #status
predicted #label MAT\
#domain extracellular #label EXT\
#domain cadherin repeat homology #label CR1\
#domain cadherin repeat homology #label CR2\
#domain cadherin repeat homology #label CR3\
#domain cadherin repeat homology #label CR4\
#domain cadherin repeat homology #label CR5\
#domain cadherin repeat homology #label CR6\
#domain cadherin repeat homology #label CR7\
#domain cadherin repeat homology #label CR8\
#domain cadherin repeat homology #label CR9\
#domain cadherin repeat homology #label CR10\
#domain cadherin repeat homology #label CR11\
#domain cadherin repeat homology #label CR12\
#domain cadherin repeat homology #label CR13\
#domain cadherin repeat homology #label CR14\
#domain cadherin repeat homology #label CR15\
#domain cadherin repeat homology #label CR16\
#domain cadherin repeat homology #label CR17\
#domain cadherin repeat homology #label CR18\
#domain cadherin repeat homology #label CR19\
#domain cadherin repeat homology #label CR20\
#domain cadherin repeat homology #label CR21\
#domain cadherin repeat homology #label CR22\
#domain cadherin repeat homology #label CR23\
#domain cadherin repeat homology #label CR24\
#domain cadherin repeat homology #label CR25\
#domain cadherin repeat homology #label CR26\
#domain cadherin repeat homology #label CR27\
#domain cadherin repeat homology #label CR28\
#domain cadherin repeat homology #label CR29\
#domain cadherin repeat homology #label CR30\
#domain cadherin repeat homology #label CR31\
#domain cadherin repeat homology #label CR32\
#domain cadherin repeat homology #label CR33\
#domain EGF homology #label EGF1\
#domain EGF homology #label EGF2\
#domain EGF homology #label EGF3\
#domain EGF homology #label EGF4\
#domain transmembrane #label TM\
#domain intracellular #label INT\
#length 5147 #molecular-weight 564895 #checksum 6994
SUMMARY
SEQUENCE
```

```
Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
```

```
X X
CCTXXX
|||
SYTLNITASDCGTPSLSTTVLYNVLVV
1460 X 1470 1480
```

2. US-08-121-713B-3 (1-7)

LDL receptor-related protein precursor - human

```
S02392
ENTRY S02392 #type complete
TITLE LDL receptor-related protein precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
28-Oct-1994
ACCESSIONS S02392; S30027; S12538
REFERENCE S02392
#authors Herz, J.; Hamann, U.; Rogne, S.; Myklebost, O.; Gausepohl,
H.; Stanley, K.K.
#journal EMBO J. (1988) 7:4119-4127
#title Surface location and high affinity for calcium of a 500-kd
liver membrane protein closely related to the LDL-receptor
suggest a physiological role as lipoprotein receptor.
#cross-references MUID:89210795
```

```
1279-1384 #domain cadherin repeat homology #label C12\
1385-1489 #domain cadherin repeat homology #label C13\
1490-1601 #domain cadherin repeat homology #label C14\
1605-1713 #domain cadherin repeat homology #label C15\
1715-1823 #domain cadherin repeat homology #label C16\
1824-1922 #domain cadherin repeat homology #label C17\
1923-2027 #domain cadherin repeat homology #label C18\
2028-2167 #domain cadherin repeat homology #label C19\
2168-2278 #domain cadherin repeat homology #label C20\
2279-2384 #domain cadherin repeat homology #label C21\
2387-2491 #domain cadherin repeat homology #label C22\
2492-2596 #domain cadherin repeat homology #label C23\
2597-2703 #domain cadherin repeat homology #label C24\
2705-2810 #domain cadherin repeat homology #label C25\
2811-2913 #domain cadherin repeat homology #label C26\
2914-3013 #domain cadherin repeat homology #label C27\
3014-3124 #domain cadherin repeat homology #label C28\
3125-3229 #domain cadherin repeat homology #label C29\
3230-3334 #domain cadherin repeat homology #label C30\
3335-3439 #domain cadherin repeat homology #label C31\
3440-3545 #domain cadherin repeat homology #label C32\
3546-3651 #domain cadherin repeat homology #label C33\
3652-3757 #domain EGF homology #label EGF1\
3758-3863 #domain EGF homology #label EGF2\
3864-3969 #domain EGF homology #label EGF3\
3970-4075 #domain EGF homology #label EGF4\
4076-4181 #domain transmembrane #label TM\
4182-4287 #domain intracellular #label INT\
4288-4393 #length 5147 #molecular-weight 564895 #checksum 6994
SUMMARY
SEQUENCE
```

#accession S02392
##molecule_type mRNA
##residues 1-4544 ##label HER
##cross-references EMBL:X13916
##note the nucleotide sequence is not given in this paper
REFERENCE
#authors Kristensen, T.
#submission submitted to the EMBL Data Library, October 1990
#accession S30027
##molecule_type mRNA
##residues 3275-3864 ##label KRI
##cross-references EMBL:X55077
REFERENCE
#authors Herz, J.; Kowal, R.C.; Goldstein, J.L.; Brown, M.S.
#journal EMBO J. (1990) 9:1769-1776
#title Proteolytic processing of the 600 kd low density lipoprotein receptor-related protein (LRP) occurs in a trans-Golgi compartment.
#cross-references MVID:90269210
#contents annotation; site of proteolytic cleavage
CLASSIFICATION
#superfamily LDL receptor ligand-binding repeat homology; LDL receptor/EGF precursor homology; LDL receptor
#YWD-containing repeat homology
#calcium binding; glycoprotein; heterodimer; transmembrane protein
KEYWORDS
#domain signal sequence #status predicted #label SIG\
#product LDL receptor-related protein 515K chain #status predicted #label 515K\
#domain LDL receptor ligand-binding repeat homology #label LDL1\
#domain LDL receptor ligand-binding repeat homology #label LDL2\
#domain LDL receptor YWTD-containing repeat homology #label YW1\
#domain LDL receptor YWTD-containing repeat homology #label YW2\
#domain LDL receptor YWTD-containing repeat homology #label YW3\
#domain LDL receptor YWTD-containing repeat homology #label YW4\
#domain LDL receptor YWTD-containing repeat homology #label YW5\
#domain LDL receptor YWTD-containing repeat homology #label YW6\
#domain LDL receptor YWTD-containing repeat homology #label YW7\
#domain LDL receptor YWTD-containing repeat homology #label YW8\
#domain LDL receptor ligand-binding repeat homology #label LDL3\
#domain LDL receptor ligand-binding repeat homology #label LDL4\
#domain LDL receptor ligand-binding repeat homology #label LDL5\
#domain LDL receptor ligand-binding repeat homology #label LDL6\
#domain LDL receptor ligand-binding repeat homology #label LDL7\
#domain LDL receptor ligand-binding repeat homology

#label LDL8\
#domain LDL receptor ligand-binding repeat homology #label LDL9\
#domain LDL receptor ligand-binding repeat homology #label LDLA\
#domain LDL receptor YWTD-containing repeat homology #label YW9\
#domain LDL receptor YWTD-containing repeat homology #label YWA\
#domain LDL receptor YWTD-containing repeat homology #label YWB\
#domain LDL receptor YWTD-containing repeat homology #label YWC\
#domain LDL receptor YWTD-containing repeat homology #label YWD\
#domain LDL receptor YWTD-containing repeat homology #label YWE\
#domain LDL receptor YWTD-containing repeat homology #label YWF\
#domain LDL receptor YWTD-containing repeat homology #label YWG\
#domain LDL receptor YWTD-containing repeat homology #label YWH\
#domain LDL receptor YWTD-containing repeat homology #label YWJ\
#domain LDL receptor YWTD-containing repeat homology #label YWK\
#domain LDL receptor YWTD-containing repeat homology #label YWL\
#domain LDL receptor YWTD-containing repeat homology #label YWM\
#domain LDL receptor YWTD-containing repeat homology #label YWN\
#domain LDL receptor YWTD-containing repeat homology #label YWO\
#domain LDL receptor YWTD-containing repeat homology #label YWP\
#domain LDL receptor YWTD-containing repeat homology #label YWQ\
#domain LDL receptor YWTD-containing repeat homology #label YWR\
#domain LDL receptor YWTD-containing repeat homology #label YWS\
#domain LDL receptor YWTD-containing repeat homology #label YWT\
#domain LDL receptor YWTD-containing repeat homology #label YWU\
#domain LDL receptor ligand-binding repeat homology #label LDLB\
#domain LDL receptor ligand-binding repeat homology #label LDLC\
#domain LDL receptor ligand-binding repeat homology #label LDLD\
#domain LDL receptor ligand-binding repeat homology #label LDLE\
#domain LDL receptor ligand-binding repeat homology #label LDLF\
#domain LDL receptor ligand-binding repeat homology #label LDLG\
#domain LDL receptor ligand-binding repeat homology

```
2819-2853 #label LDLH\
#domain LDL receptor ligand-binding repeat homology
2858-2897 #label LDLI\
#domain LDL receptor ligand-binding repeat homology
2904-2939 #label LDLJ\
#domain LDL receptor ligand-binding repeat homology
3069-3113 #label LDLK\
#domain LDL receptor YWTD-containing repeat homology
3114-3156 #label YWV\
#domain LDL receptor YWTD-containing repeat homology
3157-3200 #label YW\
#domain LDL receptor YWTD-containing repeat homology
3201-3241 #label YW\
#domain LDL receptor YWTD-containing repeat homology
3242-3283 #label YW\
#domain LDL receptor YWTD-containing repeat homology
3334-3369 #label YWZ\
#domain LDL receptor ligand-binding repeat homology
3374-3408 #label LDL\
#domain LDL receptor ligand-binding repeat homology
3413-3448 #label LDL\
#domain LDL receptor ligand-binding repeat homology
3453-3489 #label LDL\
#domain LDL receptor ligand-binding repeat homology
3494-3531 #label LDL\
#domain LDL receptor ligand-binding repeat homology
3536-3570 #label LDL\
#domain LDL receptor ligand-binding repeat homology
3575-3609 #label LDL\
#domain LDL receptor ligand-binding repeat homology
3613-3647 #label LDL\
#domain LDL receptor ligand-binding repeat homology
3654-3690 #label LDL\
#domain LDL receptor ligand-binding repeat homology
3695-3731 #label LDL\
#domain LDL receptor ligand-binding repeat homology
3741-3776 #label LDL\
#domain LDL receptor ligand-binding repeat homology
3912-3969 #label LDL\
#domain LDL receptor YWTD-containing repeat homology
3944-4544 #product LDL receptor-related protein 85K chain #status
#predicted #label 85K\
#domain extracellular #status predicted #label EXT\
3944-4420 #domain LDL receptor YWTD-containing repeat homology
3970-4012 #label YWZ\
#domain LDL receptor YWTD-containing repeat homology
4013-4056 #label YW\
#domain LDL receptor YWTD-containing repeat homology
4057-4100 #label YWZ\
#domain transmembrane #status predicted #label TM\
4421-4444 #domain intracellular #status predicted #label INT\
4445-4544 #modified site erythro-beta-hydroxyasparagine (Asn)
166,2998 #status predicted\
#modified site erythro-beta-hydroxyaspartic acid (Asp)
2958 #status predicted\
#binding site carbohydrate (Asn) (covalent) #status
4075,4125,4278 #length 4544 #molecular-weight 504571 #checksum 662
SUMMARY
SEQUENCE
```

```
Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
|||
DCMDGDEEACGTGVRTCPDEFQCNN
3640 3650 X 3660

3. US-08-121-713B-3 (1-7)
S25021 probable polyketide synthase - Bacillus subtilis

ENTRY S25021 #type complete
TITLE probable polyketide synthase - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993
ACCESSIONS S25021
REFERENCE S25021
#authors Scotti, C.; Piatti, M.; Cuzzoni, A.; Tognoni, A.; Grandi, G.;
Galizzi, A.; Albertini, A.M.
#submission submitted to the EMBL Data Library, July 1992
#description A Bacillus subtilis large ORF coding for a polypeptide highly
similar to polyketide synthases.
#accession S25021
#status Preliminary
#residues 1-4427 #label SCO
#cross-references EMBL:Z14098
SUMMARY #length 4427 #molecular-weight 493405 #checksum 2187
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
|||
HVAGAIIDLGCGTSHESMGWINLQK
3440 X 3450 3460

4. US-08-121-713B-3 (1-7)
S28600 hypothetical protein a - human coronavirus

ENTRY S28600 #type complete
TITLE hypothetical protein a - human coronavirus
ORGANISM #formal_name human coronavirus
DATE 25-Feb-1994; #sequence_revision 25-Feb-1994; #text_change
25-Feb-1994
ACCESSIONS S28600
REFERENCE S28600
#authors Herold, J.; Raabe, T.; Schelle-Prinz, B.; Siddell, S.G.
#submission submitted to the EMBL Data Library, December 1992
#accession S28600
#status Preliminary
#residues 1-4085 #label HER
```

```
##cross-references EMBL:X69721
SUMMARY #length 4085 #molecular-weight 454209 #checksum 8951
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||
NDRGVANGYVCGTGLNVLNVLNLSMFS
2720 2730 X 2740

5. US-08-121-713B-3 (1-7)
A40701 tenascin-X precursor - human

ENTRY #type complete
TITLE tenascin-X precursor - human
ORGANISM #formal name Homo sapiens #common name man
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change
ACCESSIONS A40701; A33725
REFERENCE #journal J. Cell Biol. (1993) 122:265-278
#authors Bristow, J.; Tee, M.K.; Gitelman, S.E.; Mellon, S.H.; Miller, W.L.
#journal J. Cell Biol. (1993) 122:265-278
#title Tenascin-X: a novel extracellular matrix protein encoded by the human XB gene overlapping P450c21B.
#accession A40701
##status preliminary
##molecule_type DNA
##residues 1-3566 ##label BRI
##cross-references EMBL:X71937
REFERENCE A33725
#authors Morel, Y.; Bristow, J.; Gitelman, S.E.; Miller, W.L.
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:6582-6586
#title Transcript encoded on the opposite strand of the human steroid 21-hydroxylase/complement component C4 gene locus.
#cross-references M01D:89367293
#accession A33725
##molecule_type mRNA
##residues 2748-3199, 'V', 3201-3298, 'E', 3299-3314, 'G', 3316-3566
##cross-references GB:M25813

GENETICS
#gene XB
#keywords extracellular matrix; glycoprotein; heptad repeat
SUMMARY #length 3566 #molecular-weight 385618 #checksum 7433
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||
MCWPGYTGDCGTGACPGDCGRGRVCV
```

```
490 X 500

6. US-08-121-713B-3 (1-7)
A46112 genome polyprotein - rice tungro spherical virus (
ENTRY #type complete
TITLE genome polyprotein - rice tungro spherical virus (strain Los Banos)
ORGANISM #formal name rice tungro spherical virus
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
ACCESSIONS A46112
REFERENCE #authors Shen, P.; Kaniewska, M.; Smith, C.; Beachy, R.N.
#journal Virology (1993) 193:621-630
#title Nucleotide sequence and genomic organization of rice tungro spherical virus.
#accession A46112
##molecule_type genomic RNA
##residues 1-3473 ##label SHE
##cross-references GB:S57835
CLASSIFICATION #superfamily rice tungro spherical virus genome polyprotein
KEYWORDS #polyprotein
SUMMARY #length 3473 #molecular-weight 390260 #checksum 8842
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||
VFRTFGLRLCGTYVCMPPAHYLDLITS
2670 2680

7. US-08-121-713B-3 (1-7)
S27927 polyprotein - rice tungro bacilliform virus

ENTRY #type complete
TITLE polyprotein - rice tungro bacilliform virus
ORGANISM #formal name rice tungro bacilliform virus, RTBV
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
ACCESSIONS S27927
REFERENCE #authors Shen, P.; Kaniewska, M.; Smith, C.; Beachy, R.N.
#journal submitted to the EMBL Data Library, June 1992
#accession S27927
##status preliminary
##residues 1-3473 ##label SHE
##cross-references EMBL:M95497
SUMMARY #length 3473 #molecular-weight 390260 #checksum 8842
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0
```

X X
CGTXXXX
|||
VFRFGVRLCGTYVCMAPHYLDLDEITS
2670 2680

8. US-08-121-713B-3 (1-7)
A42551 genome polyprotein - dengue virus type 1 (strain S
S275/90)
ENTRY #type complete
TITLE genome polyprotein - dengue virus type 1 (strain Singapore
S275/90)
CONTAINS capsid protein; envelope protein; membrane protein;
nonstructural protein NS1; nonstructural protein NS2a;
nonstructural protein NS2b; nonstructural protein NS3;
nonstructural protein NS4a; nonstructural protein NS4b;
nonstructural protein NS5
ORGANISM #formal name dengue virus type 1
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
09-Sep-1994
ACCESSION A42551
REFERENCE Fu, J.; Tan, B.H.; Yap, E.H.; Chan, Y.C.; Tan, Y.H.
#authors Virology (1992) 188:953-958
#journal Full-length cDNA sequence of dengue type 1 virus (Singapore
#title strain S275/90).
#accession A42551
##molecule type genomic RNA
##residues 1-3396 ##label FUJ
##cross-references GB:M87512
CLASSIFICATION #superfamily yellow fever virus genome polyprotein
KEYWORDS capsid protein; envelope protein; glycoprotein; nonstructural
protein; polyprotein; transmembrane protein
FEATURE
1-114 #product capsid protein #status predicted #label CAP\
115-281 #product membrane protein precursor #status predicted
#label MEP\
#domain signal sequence #status predicted #label SIG\
205-281 #product membrane protein #status predicted #label MEM\
267-279 #domain transmembrane #status predicted #label TM1\
282-774 #product envelope protein #status predicted #label ENV\
753-769 #domain transmembrane #status predicted #label TM2\
775-1127 #product nonstructural protein NS1 #status predicted
#label NS1\
1128-1344 #product nonstructural protein NS2a #status predicted
#label N2A\
1345-1474 #product nonstructural protein NS2b #status predicted
#label N2B\
1475-2093 #product nonstructural protein NS3 #status predicted
#label NS3\
2094-2243 #product nonstructural protein NS4a #status predicted
#label N4A\
2244-2492 #product nonstructural protein NS4b #status predicted
#label N4B\
2493-3396 #product nonstructural protein NS5 #status predicted
#label NS5\
183,347,433 #binding site carbohydrate (Asn) (covalent) #status
predicted

SUMMARY #length 3396 #molecular-weight 379561 #checksum 3242
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||
NSTHEMYWVSGGTGNIVSAVNMTSRML
2710 X 2720 2730

9. US-08-121-713B-3 (1-7)
GNWVD3 genome polyprotein - dengue virus type 3

ENTRY GNWVD3 #type complete
TITLE genome polyprotein - dengue virus type 3
CONTAINS capsid protein; envelope protein; membrane protein;
nonstructural protein NS1; nonstructural protein NS2a;
nonstructural protein NS2b; nonstructural protein NS3;
nonstructural protein NS4a; nonstructural protein NS4b;
nonstructural protein NS5
ORGANISM #formal name dengue virus type 3
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
09-Sep-1994
ACCESSION A34774
REFERENCE Osatomi, K.; Sumiyoshi, H.
#authors Virology (1990) 176:643-647
#journal Complete nucleotide sequence of dengue type 3 virus genome
#title RNA.
#cross-references MUID:90266483
#accession A34774
##molecule type genomic RNA
##residues 1-3390 ##label OSA
CLASSIFICATION #superfamily yellow fever virus genome polyprotein
KEYWORDS capsid protein; envelope protein; glycoprotein; nonstructural
protein; polyprotein; transmembrane protein
FEATURE
1-114 #product capsid protein #status predicted #label CAP\
46-67 #domain transmembrane #status predicted #label TM1\
115-280 #product membrane protein precursor #status predicted
#label MEP\
#domain signal sequence #status predicted #label SIG\
115-205 #product membrane protein #status predicted #label MEM\
206-280 #domain transmembrane #status predicted #label TM3\
266-280 #product envelope protein #status predicted #label ENV\
724-746 #domain transmembrane #status predicted #label TM4\
753-771 #domain transmembrane #status predicted #label TM5\
774-1184 #product nonstructural protein NS1 #status predicted
#label NS1\
1156-1175 #domain transmembrane #status predicted #label TM6\
1185-1343 #product nonstructural protein NS2a #status predicted
#label N2A\
1344-1473 #product nonstructural protein NS2b #status predicted
#label N2B\
1474-2092 #product nonstructural protein NS3 #status predicted
#label NS3\
183,347,433

2093-2378 #product nonstructural protein NS4a #status predicted
#label N4A\
2379-2490 #product nonstructural protein NS4b #status predicted
#label N4B\
2491-3390 #product nonstructural protein NS5 #status predicted
#label NS5\
183,347,433,750,
903,980,1132,1188,
1661,2300,2304,
2386,2456,2702,
2712 #binding site carbohydrate (Aan) (covalent) #status
predicted

SEQUENCE
#length 3390 #molecular-weight 378061 #checksum 1535

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||

EGTTVISENCGRPSLRITTVSGKL
1060 X 1070 1080

10. US-08-121-713B-3 (1-7)

macrogolgin - human

ENTRY S37536 #type complete
TITLE macrogolgin - human
ORGANISM #formal name Homo sapiens #common name man
DATE 09-Dec-1993; #sequence_revision 09-Dec-1993; #text_change
09-Dec-1993

ACCESSIONS S37536
REFERENCE S37536
#authors Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.;
Griffiths, G.; Renz, M.
#submission submitted to the EMBL Data Library, September 1993
#accession S37536

##status preliminary
##residues 1-3259 ##label SEE
##cross-references EMBL:X75304

SUMMARY #length 3259 #molecular-weight 376075 #checksum 4495
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||

DSRPTIIGSCGTQEQALLIDLFNSC
3190 3200 X 3210

11. US-08-121-713B-3 (1-7)

hypothetical protein DGF-1 - Trypanosoma cruzi

ENTRY S27852 #type complete
TITLE hypothetical protein DGF-1 - Trypanosoma cruzi
ORGANISM #formal name Trypanosoma cruzi
DATE 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
30-Sep-1993

ACCESSIONS S27852
REFERENCE S27852
#authors Wincker, P.; Murta-Dovalles, A.C.; Goldenberg, S.
#submission submitted to the EMBL Data Library, April 1992
#description Nucleotide sequence of a representative member of a
Trypanosoma cruzi dispersed gene family.

#accession S27852

##molecule_type DNA

##residues 1-3229 ##label WIN

##cross-references EMBL:M90534

SUMMARY #length 3229 #molecular-weight 334929 #checksum 5768

SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||

LFPGDVVVFCGTCNCDDAACYPGTEL
1020 1030

12. US-08-121-713B-3 (1-7)

putative cell-surface protein (cysteine-rich repea

A48450

ENTRY A48450 #type complete
TITLE putative cell-surface protein (cysteine-rich repeat motif) -
trypanosoma cruzi

ORGANISM #formal name Trypanosoma cruzi

DATE 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
18-Nov-1994

ACCESSIONS A48450

REFERENCE A48450

#authors Wincker, P.; Murto-Dovalles, A.C.; Goldenberg, S.

#journal Mol. Biochem. Parasitol. (1992) 55:217-220

#title Nucleotide sequence of a representative member of a

Trypanosoma cruzi dispersed gene family.

#cross-references MIMD:93063053

#accession A48450

##status preliminary

##molecule_type DNA

##residues 1-3229 ##label WIN

##cross-references NCBI:P118407

##note sequence extracted from NCBI backbone

##note sequence not compared to nucleotide translation

SUMMARY #length 3229 #molecular-weight 334929 #checksum 5768

SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X


```
CGTXXXX
||||
LFPGDVVVFSGTCNDDAACYPGTEL
1020      1030

13. US-08-121-713B-3 (1-7)
A44062 genome polyprotein - pepper mottle virus (strain C)
ENTRY      A44062 #type complete
TITLE      genome polyprotein - pepper mottle virus (strain California)
CONTAINS   29K protein; 34K protein; coat protein; cylindrical inclusion
           protein; helper component protein; nuclear inclusion
           protein a; nuclear inclusion protein b
ORGANISM   #formal name pepper mottle virus
DATE       31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
ACCESSIONS A44062
REFERENCE   A44062
#authors   Vance, V.B.; Moore, D.; Turpen, T.H.; Bracker, A.; Hollowell,
#journal   V.C.
#title     Virology (1992) 191:19-30
#keywords  The complete nucleotide sequence of pepper mottle virus
           genomic RNA: comparison of the encoded polyprotein with
           those of other sequenced potyviruses.
#accession A44062
#molecule_type genomic RNA
#residues  1-3068 #label VAN
#cross-references GB:M96425
CLASSIFICATION #superfamily tobacco etch virus genome polyprotein
               coat protein; cylindrical inclusion protein; inclusion
               protein; nucleus; polyprotein
FEATURE
1-287      #product 34K protein #status predicted #label KPT\
288-743    #product helper component protein #status predicted
           #label HCP\
744-1156   #product 29K protein #status predicted #label KPR\
1157-1790  #product cylindrical inclusion protein #status predicted
           #label CIP\
1791-2276  #product nuclear inclusion protein a #status predicted
           #label NIA\
2277-2799  #product nuclear inclusion protein b #status predicted
           #label NIB\
2800-3068  #product coat protein #status predicted #label CPT
SUMMARY    #length 3068 #molecular-weight 348653 #checksum 2964
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
||||
YFADAEFEFCGYEVYRHQSSRSRTL
2780      2790 X 2800

14. US-08-121-713B-3 (1-7)
JQ1917 polyprotein - parsnip yellow fleck virus
```

```
JQ1917 #type complete
ENTRY  polyprotein - parsnip yellow fleck virus
TITLE  #formal name parsnip yellow fleck virus, PVFV
ORGANISM
DATE    31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
ACCESSIONS JQ1917
REFERENCE   JQ1917
#authors   Turnbull-Ross, A.D.; Reavy, B.; Mayo, M.A.; Murrant, A.F.
#journal   J. Gen. Virol. (1992) 73:3203-3211
#title     The nucleotide sequence of parsnip yellow fleck virus: a
           plant picorna-like virus.
#contents  Strain P-121
#accession JQ1917
#molecule_type mRNA
#residues  1-3027 #label TUR
KEYWORDS  polyprotein
SUMMARY   #length 3027 #molecular-weight 336244 #checksum 9823
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
||||
YIFSTTVPGGCTKHGLADPGAFMRRR
1570 X 1580

15. US-08-121-713B-3 (1-7)
B39658 polyprotein coli protein DP2.5 - human
ENTRY      B39658 #type complete
TITLE      polyprotein coli protein DP2.5 - human
ORGANISM   #formal name Homo sapiens #common name man
DATE       21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change
ACCESSIONS B39658
REFERENCE   A39658
#authors   Joslyn, G.; Carlson, M.; Thliveris, A.; Albertsen, H.;
           Gelbert, L.; Samowitz, W.; Groden, J.; Stevens, J.; Spirio,
           L.; Robertson, M.; Sargeant, L.; Krapcho, K.; Wolff, E.;
           Burt, R.; Hughes, J.P.; Warrington, J.; McPherson, J.;
           Wasmuth, J.; Le Paslier, D.; Abderrahim, H.; Cohen, D.;
           Leppert, M.; White, R.
#journal   Cell (1991) 66:601-613
#title     Identification of deletion mutations and three new genes at
           the familial polyposis locus.
#cross-references MUID:91330307
#accession B39658
#status    preliminary
#molecule_type DNA
#residues  1-2844 #label JOS
#cross-references GB:M73548
CLASSIFICATION #superfamily adenomatous polyposis coli protein
SUMMARY       #length 2844 #molecular-weight 311654 #checksum 7518
SEQUENCE
```

Initial Score = 7 Optimized Score = 7 Significance = 1.84
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||
SHSLTIVSNACGTLWNLSARNPKDQEA
680 690

> O < IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq3sept.res made by on Fri 19 May 95 8:54:49-PDT.

Query sequence being compared: US-08-121-713B-3 (1-7)
Number of sequences searched: 43470
Number of scores above cutoff: 4588

Results of the initial comparison of US-08-121-713B-3 (1-7) with:
Data bank : Swiss-Prot 31, all entries

100000-
N -
U50000-
M -
B -
E -
R -
O -
F10000-
S -
E 5000-
Q -
U -
E -
N -
C -
S 1000-
E -
S 500-
100-
50-
*
*
*

-
-
-
-
-
-
10-
-
-
5-
-
-
-
-
-
-
0-

SCORE 0 1 1 1 2 2 3 4 5 6 7
STDEV 1 1 1 1 1 1 1 1 1 1 1

PARAMETERS

Similarity matrix Unitary K-tuple 2
Mismatch penalty 1 Joining penalty 20
Gap penalty 1.00 Window size 6
Gap size penalty 0.05
Cutoff score 0
Randomization group 0
Initial scores to save 45 Alignments to save 15
Optimized scores to save 0 Display context 10

SEARCH STATISTICS

Scores: Mean 3 Median 6 Standard Deviation 2.02
Times: CPU 00:00:44.00 Total Elapsed 00:00:46.00

Number of residues: 15335248
Number of sequences searched: 43470
Number of scores above cutoff: 4588

Cut-off raised to 5.
Cut-off raised to 6.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

1019 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Opt. Score	Sig. Frame
1. FAT DROME	CADHERIN-RELATED TUMOR SUPPRE	5147	7	7 1.98 0
2. PKSI_BACSU	POTATIVE POLYKETIDE SYNTHASE	4427	7	7 1.98 0

3. RPOA_CVH22 RNA-DIRECTED RNA POLYMERASE (4085 7 7 1.98 0
4. RPOA_LELV RNA-DIRECTED RNA POLYMERASE (3859 7 7 1.98 0
5. POLG_DEN1S GENOME POLYPROTEIN (CONTAINS: 3396 7 7 1.98 0
6. POLG_DEN3 GENOME POLYPROTEIN (CONTAINS: 3390 7 7 1.98 0
7. POLG_PEMVC GENOME POLYPROTEIN (CONTAINS: 3068 7 7 1.98 0
8. POLG_PPV1 GENOME POLYPROTEIN (CONTAINS: 3027 7 7 1.98 0
9. APC_HUMAN ADENOMATOUS POLYPOSIS COLI PR 2843 7 7 1.98 0
10. VWF_HUMAN VON WILLEBRAND FACTOR PRECURS 2813 7 7 1.98 0
11. IP3R_MOUSE INOSITOL 1,4,5-TRISPHOSPHATE- 2749 7 7 1.98 0
12. IP3R_RAT INOSITOL 1,4,5-TRISPHOSPHATE- 2749 7 7 1.98 0
13. ZEP1_HUMAN ZINC FINGER PROTEIN 40 (HUMAN 2717 7 7 1.98 0
14. G156_PAPR 156G SURFACE PROTEIN PRECURSO 2715 7 7 1.98 0
15. CENP_HUMAN CENTROMERIC PROTEIN E (CENP-E 2663 7 7 1.98 0
16. RRPB_HVBV DNA-DIRECTED RNA POLYMERASE (2652 7 7 1.98 0
17. MPRI_BOVIN CATION-INDEPENDENT MANNOSE-6- 2499 7 7 1.98 0
18. MPRI_HUMAN CATION-INDEPENDENT MANNOSE-6- 2491 7 7 1.98 0
19. F1NC_RAT FIBRONECTIN PRECURSOR (FN) . 2477 7 7 1.98 0
20. RPB1_P1AFD DNA-DIRECTED RNA POLYMERASE I 2452 7 7 1.98 0
21. CCB1_RABIT BRAIN CALCIUM CHANNEL BI-2 PR 2424 7 7 1.98 0
22. F1NC_HUMAN FIBRONECTIN PRECURSOR. 2386 7 7 1.98 0
23. YIM9_YEAST HYPOTHETICAL 269.9 KD PROTEIN 2376 7 7 1.98 0
24. FABI_YEAST FABI PROTEIN. 2278 7 7 1.98 0
25. MUC2_HUMAN MUCIN 2 (INTESTINAL MUCIN 2) 2274 7 7 1.98 0
26. CCB1_RABIT BRAIN CALCIUM CHANNEL BI-1 PR 2273 7 7 1.98 0
27. F1NC_BOVIN FIBRONECTIN (FN) . 2265 7 7 1.98 0
28. CCB3_RABIT BRAIN CALCIUM CHANNEL BII-1 P 2259 7 7 1.98 0
29. POL1_GCMV RNA1 POLYPROTEIN (250 KD PROT 2252 7 7 1.98 0
30. GLSN_MEDSA GLUTAMATE SYNTHASE (NADH) PRE 2194 7 7 1.98 0
31. CCB4_RABIT BRAIN CALCIUM CHANNEL BII-2 P 2178 7 7 1.98 0
32. POLG_HRV1B GENOME POLYPROTEIN (COAT PROT 2157 7 7 1.98 0
33. POLG_HRV2 GENOME POLYPROTEIN (COAT PROT 2150 7 7 1.98 0
34. CRB_DROME CRUMBS PROTEIN PRECURSOR (95F 2139 7 7 1.98 0
35. AGR1_CHICK AGRIN PRECURSOR. 1955 7 7 1.98 0
36. RPOA_ACLSIV RNA-DIRECTED RNA POLYMERASE (1884 7 7 1.98 0
37. C1C1_RABIT DIHYDROPYRIDINE-SENSITIVE L- 1873 7 7 1.98 0
38. BRC1_HUMAN BREAST CANCER TYPE 1 SUSCEPTI 1863 7 7 1.98 0
39. RPB1_CAEEL DNA-DIRECTED RNA POLYMERASE I 1859 7 7 1.98 0
40. POLR_EPMV RNA REPLICASE POLYPROTEIN (EC 1839 7 7 1.98 0
41. LMB1_HUMAN LAMININ BETA-1 CHAIN PRECURSO 1786 7 7 1.98 0
42. LMB1_MOUSE LAMININ BETA-1 CHAIN PRECURSO 1786 7 7 1.98 0
43. RPOA_SHVX RNA REPLICATING PROTEIN (CONT 1718 7 7 1.98 0
44. TCFB_RAT TRANSFORMING GROWTH FACTOR BE 1712 7 7 1.98 0
45. RPOC_THEME DNA-DIRECTED RNA POLYMERASE B 1690 7 7 1.98 0

1. US-08-121-713B-3 (1-7)
FAT_DROME CADHERIN-RELATED TUMOR SUPPRESSOR PRECURSOR (FAT P

ID FAT_DROME STANDARD; PRT; 5147 AA.

AC P33450;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE CADHERIN-RELATED TUMOR SUPPRESSOR PRECURSOR (FAT PROTEIN).
GN FT.

OS DROSOPHILA MELANOGASTER (FRUIT FLY).

OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.

RN [1]

RM SEQUENCE FROM N.A.

RP 92069752

RA MAHONEY P.A., WEBER U., ONOFRECHUK P., BIESSMANN H., BRYANT P.J.,
RA GOODMAN C.S.;
RL CELL 67:853-868(1991).
CC -1- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.
CC -1- DISEASE: RECESSIVE LETHAL MUTATIONS IN FAT CAUSE HYPERELASTIC,
CC TUMOR-LIKE OVERGROWTH OF LARVAL IMAGINAL DISCS, DEFECTS IN
CC DIFFERENTIATION AND MORPHOGENESIS, AND DEATH DURING THE POPAL
CC STAGE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY OF CELL ADHESION
CC MOLECULES. CONTAINS 37 CADHERINS-TYPE REPEATS.
CC -1- SIMILARITY: THE PROTEIN INCLUDES 4 EGF-LIKE REPEATS.
DR EMBL; M80537; DMFAT.
DR PIR; A41087; IJFFTM.
DR FLYBASE; FBGN0001075; FT.
DR PROSITE; PS00022; EGF.
DR PROSITE; PS00232; CADHERIN.
KW CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
KW CALCIUM-BINDING; REPEAT; EGF-LIKE DOMAIN;
FT SIGNAL 1 35
FT CHAIN 36 5147
FT DOMAIN 36 4583
FT TRANSSEM 4584 4609
FT DOMAIN 4610 5147
FT REPEAT 157 270
FT REPEAT 271 382
FT REPEAT 383 494
FT REPEAT 495 599
FT REPEAT 600 708
FT REPEAT 709 820
FT REPEAT 821 942
FT REPEAT 943 1049
FT REPEAT 1050 1153
FT REPEAT 1154 1278
FT REPEAT 1279 1384
FT REPEAT 1385 1489
FT REPEAT 1490 1601
FT REPEAT 1602 1713
FT REPEAT 1714 1823
FT REPEAT 1824 1922
FT REPEAT 1923 2027
FT REPEAT 2028 2167
FT REPEAT 2168 2278
FT REPEAT 2279 2385
FT REPEAT 2386 2491
FT REPEAT 2492 2596
FT REPEAT 2597 2703
FT REPEAT 2704 2810
FT REPEAT 2811 2913
FT REPEAT 2914 3013
FT REPEAT 3014 3124
FT REPEAT 3125 3229
FT REPEAT 3230 3334
FT REPEAT 3335 3439
FT REPEAT 3440 3545
FT REPEAT 3546 3651
FT REPEAT 3652 3756
FT REPEAT 3954 4016
FT REPEAT 4017 4055
EGF-LIKE 1.
EGF-LIKE 2.

FT REPEAT 4056 4095 EGF-LIKE 3.
FT REPEAT 4096 4134 EGF-LIKE 4.
FT CARBOHYD 239 239 POTENTIAL.
FT CARBOHYD 257 257 POTENTIAL.
FT CARBOHYD 276 276 POTENTIAL.
FT CARBOHYD 280 280 POTENTIAL.
FT CARBOHYD 402 402 POTENTIAL.
FT CARBOHYD 461 461 POTENTIAL.
FT CARBOHYD 605 605 POTENTIAL.
FT CARBOHYD 631 631 POTENTIAL.
FT CARBOHYD 1155 1155 POTENTIAL.
FT CARBOHYD 1367 1367 POTENTIAL.
FT CARBOHYD 1458 1458 POTENTIAL.
FT CARBOHYD 1751 1751 POTENTIAL.
FT CARBOHYD 1831 1831 POTENTIAL.
FT CARBOHYD 1880 1880 POTENTIAL.
FT CARBOHYD 2080 2080 POTENTIAL.
FT CARBOHYD 2171 2171 POTENTIAL.
FT CARBOHYD 2247 2247 POTENTIAL.
FT CARBOHYD 2290 2290 POTENTIAL.
FT CARBOHYD 2437 2437 POTENTIAL.
FT CARBOHYD 2581 2581 POTENTIAL.
FT CARBOHYD 2799 2799 POTENTIAL.
FT CARBOHYD 2920 2920 POTENTIAL.
FT CARBOHYD 2946 2946 POTENTIAL.
FT CARBOHYD 2967 2967 POTENTIAL.
FT CARBOHYD 3167 3167 POTENTIAL.
FT CARBOHYD 3303 3303 POTENTIAL.
FT CARBOHYD 3386 3386 POTENTIAL.
FT CARBOHYD 3389 3389 POTENTIAL.
FT CARBOHYD 3525 3525 POTENTIAL.
FT CARBOHYD 3852 3852 POTENTIAL.
FT CARBOHYD 3865 3865 POTENTIAL.
FT CARBOHYD 3905 3905 POTENTIAL.
FT CARBOHYD 4306 4306 POTENTIAL.
FT CARBOHYD 4414 4414 POTENTIAL.
FT CARBOHYD 4471 4471 POTENTIAL.
FT CARBOHYD 4487 4487 POTENTIAL.
FT CARBOHYD 4539 4539 POTENTIAL.
FT CARBOHYD 4550 4550 POTENTIAL.
FT VARIANT 1229 1229 S -> G.
FT VARIANT 1233 1233 G -> S.
SQ SEQUENCE 5147 AA; 564885 MW; 23531637 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
|||

SYTLNITASDCGTPSLSTTVLYNLVW
1460 X 1470 1480

2. US-08-121-713B-3 (1-7)
PKSL_BACSU PUTATIVE POLYKETIDE SYNTHASE PKSL (PKS).

ID PKSL_BACSU STANDARD; PRT; 4427 AA.
AC Q05470;

DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
GN PUTATIVE POLYKETIDE SYNTHASE PKSL (PKS).
OS PKSL OR PKSX OR PKSA OR OUTG.
OC BACILLUS SUBTILIS.
RN PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / PB1424;
RM 93345824
RA SCOTTI C., PIATTI M., CUZZONI A., PERANI P., TOGNONI A., GRANDI G.,
RL GALIZZI A., ALBERTINI A.M.;
CC GENE 130:65-71(1993).
CC -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
SECONDARY METABOLISM.
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
DR EMBL; M97902; BSKPSX.
DR EMBL; U11039; BS11039.
DR PIR; S25021; S25021.
DR SUBTILIS; BG10698; PKSL.
KW TRANSFERASE; ACYLTRANSFERASE; ANTIBIOTIC BIOSYNTHESIS; NADP;
PHOSPHOPANTETHEINE; MULTIFUNCTIONAL ENZYME; REPEAT.
FT DOMAIN 184 282 ACYL CARRIER (ACP).
FT DOMAIN 382 759 BETA-KETOACYL SYNTHASE (KS).
FT DOMAIN 937 1115 DEHYDRATASE.
FT DOMAIN 1409 1602 BETA-KETOACYL REDUCTASE (KR).
FT DOMAIN 1665 1761 ACYL CARRIER (ACP).
FT DOMAIN 1876 2253 BETA-KETOACYL SYNTHASE (KS).
FT DOMAIN 2469 2560 ACYL CARRIER (ACP).
FT DOMAIN 2609 2702 ACYL CARRIER (ACP).
FT DOMAIN 2823 3182 BETA-KETOACYL SYNTHASE (KS).
FT DOMAIN 3575 3776 BETA-KETOACYL REDUCTASE (KR).
FT DOMAIN 3834 3925 ACYL CARRIER (ACP).
FT DOMAIN 4019 4373 BETA-KETOACYL SYNTHASE (KS).
FT BINDING 2523 2523 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 2664 2664 PHOSPHOPANTETHEINE (POTENTIAL).
SQ SEQUENCE 4427 AA; 493398 MW; 21754976 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
|||

HVAGAIIDLIGCGTSHESMGWNLQK
3440 X 3450 3460

3. US-08-121-713B-3 (1-7)
RPPA_CVH22 RNA-DIRECTED RNA POLYMERASE (ORF1A) (EC 2.7.7.48).

ID RPPA_CVH22 STANDARD; PRT; 4085 AA.

AC Q05002;

DT 01-FEB-1994 (REL. 28, CREATED)

DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)

DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)

DE RNA-DIRECTED RNA POLYMERASE (ORF1A) (EC 2.7.7.48).

OS HUMAN CORONAVIRUS (STRAIN 229E).
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; CORONAVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RM 93331726
RA HEROLD J., RAABE T., SCHELLE-PRINZ B., SIDDELL S.G.;
RL VIROLOGY 195:660-691(1993).
CC -1- FUNCTION: THE RNA DEPENDENT RNA POLYMERASE OF CORONAVIRUSES IS
CC A MULTIFUNCTIONAL PROTEIN: IT CONTAINS THE ACTIVITIES NECESSARY
CC FOR THE TRANSCRIPTION OF NEGATIVE STRANDED RNA, LEADER RNA,
CC SUBGENOMIC MRNAs AND PROGENY VIRION RNA.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE
CC + RNA(N).
CC EMBL; X69721; HCVORF1AB.
DR PIR; S28600; S28600.
KW RNA-DIRECTED RNA POLYMERASE.
SQ SEQUENCE 4085 AA; 454206 MW; 19863052 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||
NDGRVANGYVCGTGLNVLVFNILSMFS
2720 2730 X 2740

4. US-08-121-713B-3 (1-7)
RPOA_LELV RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (ORF1A/1)

ID RPOA_LELV STANDARD; PRT; 3859 AA.
AC Q04561;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (ORF1A/1B).
GN POL.
OS LELYSTAD VIRUS (LV) (PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME
OS VIRUS).
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; TOGAVIRIDAE;
OC ARTERIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RM 93297139
RA MEULENBERG J.J.M., HULST M.M., DE MEIJER E.J., MOONEN P.L.J.M.,
RA DEN BESTEN A., DE KLUYVER E.P., WENSVOORT G., MOORMANN R.J.M.;
RL VIROLOGY 192:62-72(1993).
RN [2]
RP SEQUENCE OF 3328-3859 FROM N.A.
RC STRAIN=BOXMEER 10;
RM 93174942
RA CONZELMANN K.K., VISSER N., VAN WOENSEL P., THIEL H.J.;
RL VIROLOGY 193:329-339(1993).
CC -1- FUNCTION: RNA-DIRECTED RNA POLYMERASE & POSSIBLE HELICASE. A
CC ROLE FOR NTP-BINDING PROTEINS IN RNA DUPLEX UNWINDING HAS BEEN
CC SUGGESTED. POSSIBLE CONTAINS A PROTEASE DOMAIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + RNA(N).

CC -1- SIMILARITY: WITH THE POLYMERASE PROTEIN OF OTHER CORONAVIRUSES AND
CC OF TOROVIRUSES.
DR EMBL; M96262; LEYPOLYEN.
DR EMBL; L04493; PRWPOLGLY.
DR PIR; A36861; A36861.
DR PIR; A45392; A45392.
KW RNA-DIRECTED RNA POLYMERASE; HELICASE; ATP-BINDING; HYDROLASE;
KW SERINE PROTEASE.
FT CHAIN 1 2396
FT CHAIN 2397 3859
FT CHAIN 2397 3859
FT DOMAIN 262 290
FT DOMAIN 472 525
FT DOMAIN 792 1506
FT DOMAIN 1203 1284
FT DOMAIN 1724 1830
FT DOMAIN 2010 2108
FT DOMAIN 2768 2972
FT ACT_SITE 1732 1732
FT ACT_SITE 1757 1757
FT ACT_SITE 1810 1810
FT NP_BIND 2587 2595
FT ZN_FING 3042 3094
FT CONFLICT 3506 3506
FT CONFLICT 3744 3744
SQ SEQUENCE 3859 AA; 421843 MW; 21900374 CN;

ORF1A.
ORF1B.
CYS-RICH.
CYS-RICH.
HELICASE.
CYS-RICH.
TRYPSIN-LIKE SERINE PROTEASE.
POLYMERASE.
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
ATP (BY SIMILARITY).
BY SIMILARITY.
T -> V (IN REF. 2).
V -> I (IN REF. 2).
Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||

TGPILCHVEHCGTESGDSSEPLDLSDA
810 820 X 830

5. US-08-121-713B-3 (1-7)

POLG_DENIS GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CO

ID POLG_DENIS STANDARD; PRT; 3396 AA.
AC P33478;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
DE PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED
DE RNA POLYMERASE (EC 2.7.7.48) (NS5)).
OS DENGUE VIRUS TYPE 1 (STRAIN SINGAPORE S275/90).
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;
OC FLAVIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RM 92263809
RA FU J., TAN B.H., YAP E.H., CHAN Y.C., TAN Y.H.;
RL VIROLOGY 188:953-958(1992).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF THREE PROTEINS:
CC PROTEIN PRM, PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A

CC COMPLEX OF PROTEIN C AND MRNA.
DR EMBL; M67512; DENTISEQ.
DR PIR; A42551; A42551.
KW POLYPEPTIDE; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN;
KW COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE;
KW NONSTRUCTURAL PROTEIN.
FT CHAIN 1 114
FT PROPEP 115 205
FT CHAIN 206 280
FT CHAIN 281 774
FT CHAIN 775 1127
FT CHAIN 1128 1344
FT CHAIN 1345 1474
FT CHAIN 1475 2093
FT CHAIN 2094 2243
FT CHAIN 2244 2492
FT CHAIN 2493 3396
FT NP BIND 1668 1675
FT SITE 1759 1762
FT TRANSMEM 267 279
FT TRANSMEM 753 769
FT CARBOHYD 183 183
FT CARBOHYD 347 347
FT CARBOHYD 433 433
SQ SEQUENCE 3396 AA; 379558 MW; 226867111 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||
NSTHEMYVWVSCGTGNTVSANVTSMRL
2710 X 2720 2730

6. US-08-121-713B-3 (1-7)
POLG_DEN3 GENOME POLYPEPTIDE (CONTAINS: CAPSID PROTEIN C (CO

ID POLG_DEN3 STANDARD; PRT; 3390 AA.
AC P27915;
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE GENOME POLYPEPTIDE (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E;
DE NONSTRUCTURAL PROTEINS NS1, NS2, NS4A AND NS4B; HELICASE (NS3);
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)).
OS DENGUE VIRUS TYPE 3
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;
OC FLAVIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RM 90266483
RA OSATOMI K., SUMIYOSHI H.;
RL VIROLOGY 176:643-647(1990).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
DR EMBL; M93130; DENCME.
DR PIR; A34774; GNWV03.
KW POLYPEPTIDE; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN;
KW COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE;
KW NONSTRUCTURAL PROTEIN.
FT CHAIN 1 114
FT PROPEP 115 205
FT CHAIN 206 280
FT CHAIN 281 773
FT CHAIN 774 1184
FT CHAIN 1185 1343
FT CHAIN 1344 1473
FT CHAIN 1474 2092
FT CHAIN 2093 2378
FT CHAIN 2379 2490
FT CHAIN 2491 3390
FT NP BIND 1667 1674
FT SITE 1758 1761
FT TRANSMEM 46 67
FT TRANSMEM 266 280
FT TRANSMEM 724 746
FT TRANSMEM 753 771
FT TRANSMEM 1156 1175
FT CARBOHYD 183 183
FT CARBOHYD 347 347
FT CARBOHYD 433 433
FT CARBOHYD 750 750
FT CARBOHYD 903 903
FT CARBOHYD 980 980
FT CARBOHYD 1132 1132
FT CARBOHYD 1188 1188
FT CARBOHYD 1661 1661
FT CARBOHYD 2300 2300
FT CARBOHYD 2304 2304
FT CARBOHYD 2456 2456
FT CARBOHYD 2702 2702
FT CARBOHYD 2712 2712
SQ SEQUENCE 3390 AA; 378057 MW; 22373526 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXXX
|||
EGTTVISENCGTRGPSLRITTVSGKL
1060 X 1070 1080

7. US-08-121-713B-3 (1-7)
POLG_PEMVC GENOME POLYPEPTIDE (CONTAINS: N-TERMINAL PROTEIN;

ID POLG_PEMVC STANDARD; PRT; 3068 AA.
AC Q01500;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)

DE GENOME POLYPROTEIN (CONTAINS: N-TERMINAL PROTEIN; HELPER COMPONENT-
DE PROTEINASE (HC-PRO); 42-50 KD PROTEIN; CYTOPLASMIC INCLUSION PROTEIN
DE (CI); 6 KD PROTEIN; NUCLEAR INCLUSION PROTEIN A (NI-A) (EC 3.4.-.-)
DE (49K PROTEINASE) (49 KD-PRO); NUCLEAR INCLUSION PROTEIN B (NI-B)
DE (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48); COAT PROTEIN).
OS PEPPER MOTTLE VIRUS (CALIFORNIA ISOLATE) (PEMV).
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; POTYVIRIDAE.
RN [1]
RM SEQUENCE FROM N.A.
RA VANCE V.B., MOORE D., TURPEN T.H., BRACKER A., HOLLOWELL V.C.;
RL VIROLOGY 191:19-30(1992).
CC -1- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC -1- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC MAY BE INVOLVED IN REPLICATION.
CC -1- FUNCTION: NUCLEAR INCLUSION PROTEIN A, OR THE N-TERMINAL PART OF
CC IT, VPG, IS COVALENTLY LINKED TO THE GENOMIC RNA. NI-A ALSO HAS
CC PROTEOLYTIC ACTIVITY.
CC -1- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
CC POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC INDIVIDUAL PROTEINS.
CC -1- SIMILARITY: TO OTHER POTYVIRUSES POLYPROTEINS.
CC EMBL; M96425; PMVCG.
DR PIR; A44062; A44062.
KW HYDROLASE; PROTEASE; RNA-DIRECTED RNA POLYMERASE; COAT PROTEIN;
KM POLYPROTEIN; COVALENT PROTEIN-RNA LINKAGE.
FT CHAIN 1 287 N-TERMINAL PROTEIN.
FT CHAIN 288 743 HELPER COMPONENT-PROTEINASE.
FT CHAIN 744 1156 29 KD PROTEIN.
FT CHAIN 1157 1790 CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN 1791 1842 6 KD PROTEIN.
FT CHAIN 1843 2276 NUCLEAR INCLUSION PROTEIN A.
FT CHAIN 2277 2795 NUCLEAR INCLUSION PROTEIN B.
FT CHAIN 2796 3068 COAT PROTEIN.
FT BINDING 1906 1906 COVALENT LINKAGE OF VIRAL RNA
FT (BY SIMILARITY).
SQ SEQUENCE 3068 AA; 348651 MW; 22258851 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
|||
YFADAEFEFCGTGYEVRHQSSRSDDL
2780 2790 X 2800

8. US-08-121-713B-3 (1-7)
POLG_PYPV1 GENOME POLYPROTEIN (CONTAINS: 22.5 KD PROTEIN; 26

ID POLG_PYPV1 STANDARD; PRT; 3027 AA.

AC Q05057;

DT 01-FEB-1994 (REL. 28, CREATED)

DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)

DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)

DE GENOME POLYPROTEIN (CONTAINS: 22.5 KD PROTEIN; 26 KD PROTEIN; 31 KD

DE PROTEIN; PROBABLE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48)).
OS PARSNIP YELLOW FLECK VIRUS (ISOLATE P-121) (PYFV).
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; SEQUIVIRIDAE.
RN [1]
RM SEQUENCE FROM N.A.
RA TURNBULL-ROSS A.D., REAVY B., MAYO M.A., MURANT A.F.;
RL J. GEN. VIROL. 73:3203-3211(1992).
CC -1- SIMILARITY: SOME, TO THE CMFV AND TBRV POLYPROTEINS.
DR EMBL; D14066; PYFPOLYP.
KW POLYPROTEIN; ATP-BINDING; COAT PROTEIN; RNA-DIRECTED RNA POLYMERASE.
FT BIND 1467 1474 ATP (POTENTIAL).
FT VARIANT 962 962 T -> I.
FT VARIANT 1373 1373 L -> F.
SQ SEQUENCE 3027 AA; 336242 MW; 22418420 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
|||
YIFSTNVPGCGTKRHGLADPGAFMRRR
1570 X 1580

9. US-08-121-713B-3 (1-7)
APC_HUMAN ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN).

ID APC_HUMAN STANDARD; PRT; 2843 AA.

AC P25054;

DT 01-MAY-1992 (REL. 22, CREATED)

DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)

DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)

DE ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN).

GN APC OR DP2.5.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RM 91335210

RA KINZLER K.W., NILBERT M.C., SU L.-K., VOGELSTEIN B., BRYAN T.M.,

RA LEVY D.B., SMITH K.J., PREISINGER A.C., HEDGE P., MCKECHNIE D.,

RA FINNIEAR R., MARKHAM A., GROFFEN J., BOGUSKI M.S., ALTSCHUL S.F.,

RA HORII A., ANDO H., MIYOSHI Y., MIKI Y., NISHISHO I., NAKAMURA Y.;

RL SCIENCE 253:661-665(1991).

RN [2]

RP VARIANTS FAP.

RM 91335211

RA NISHISHO I., NAKAMURA Y., MIYOSHI Y., MIKI Y., ANDO H., HORII A.,

RA KOYAMA K., UTSUNOMIYA J., BABA S., HEDGE P., MARKHAM A., KRUSH A.J.,

RA PETERSEN G., HAMILTON S.R., NILBERT M.C., LEVY D.B., BRYAN T.M.,

RA PREISINGER A.C., SMITH K.J., SU L.-K., KINZLER K.W., VOGELSTEIN B.;

RL SCIENCE 253:665-669(1991).

RN [3]

RP VARIANTS FAP.

RM 93265030

RA MIYOSHI Y., NAGASE H., ANDO H., ICHII S., NAKATSURO S., AOKI T.,
 RA MIKI Y., MORI T., NAKAMURA Y.;
 RL HUM. MOL. GENET. 1:229-233(1992).
 RN [4]
 RP VARIANTS FAP.
 RM 93244793
 RA NAKATSURO S., YANAGISAWA A., ICHII S., TAHARA E., KATO Y.,
 RA NAKAMURA Y., HORII A.;
 RL HUM. MOL. GENET. 1:559-563(1992).
 CC -1- DISEASE: MUTATIONS OF THE GENE FOR THIS PROTEIN WERE FOUND IN
 CC PATIENTS WITH FAMILIAL ADENOMATOUS POLYPOSIS (FAP) AND GARDNERS
 CC SYNDROME (GS), THAT CONTRIBUTE TO TUMOR DEVELOPMENT IN PATIENTS
 CC WITH NONINHERITED FORMS OF COLORECTAL CANCER.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES TYPES.
 CC EMBL; M74088; HSPAPAPC.
 DR PIR; A37261; RBHUAP.
 DR MIM; 175100; 11TH EDITION.
 KW ANTI-ONCOGENE, DISEASE MUTATION.
 FT DOMAIN 1 730 LEUCINE-RICH.
 FT COIL (POTENTIAL).
 FT COIL (POTENTIAL).
 FT SER-RICH.
 FT ASP/GLU-RICH (ACIDIC).
 FT ASP/GLU-RICH (ACIDIC).
 FT HIGHLY CHARGED.
 FT R -> C (IN A FAP).
 FT I -> C (IN A GASTRIC CANCER).
 FT I -> T (IN A GASTRIC CANCER).
 FT E -> G (IN A FAP).
 FT N -> D (IN A GASTRIC CANCER).
 FT G -> E (IN A GASTRIC CANCER).
 FT F -> S (IN A GASTRIC CANCER).
 FT I -> T (IN A GASTRIC CANCER).
 FT G -> E (IN A GASTRIC CANCER).
 FT T -> A (IN A FAP).
 FT V -> A (IN A GASTRIC CANCER).
 FT SEQUENCE 2843 AA; 311658 MW; 18964414 CN;
 SQ
 Initial Score = 7 Optimized Score = 7 Significance = 1.98
 Residue Identity = 42% Matches = 3 Mismatches = 4
 Gaps = 0 Conservative Substitutions = 0
 X X
 CGTXXX
 SHSLTIVSNACGTLWNLARNPKDQEA
 680 690
 10. US-08-121-713B-3 (1-7)
 VWF_HUMAN VON WILLEBRAND FACTOR PRECURSOR.
 ID VWF_HUMAN STANDARD; PRT; 2813 AA.
 AC P04275;
 DT 20-MAR-1987 (REL. 04, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE VON WILLEBRAND FACTOR PRECURSOR.
 GN F8VWF OR VWF.

OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 90462044
 RA MANCUSO D.J., TULEY E.A., WESTFIELD L.A., WORRALL N.K.,
 RA SHELTON-INLOES B.B., SORACE J.M., ALEVY Y.G., SADLER J.E.;
 RL J. BIOL. CHEM. 264:19514-19527(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RM 87016349
 RA BONTHON D., ORR E.C., MITSOCK L.M., MITSOCK L.M., GINSBURG D.,
 RA HANDIN R.I., ORKIN S.H.;
 RL NUCLEIC ACIDS RES. 14:7125-7128(1986).
 RN [3]
 RP SEQUENCE OF 1-1400 FROM N.A.
 RM 87004550
 RA VERWEIJ C.L., DIERGAARDE P.J., HART M., PANNEKOEK H.;
 RL EMBO J. 5:1839-1847(1986).
 RN [4]
 RP ERRATUM.
 RA VERWEIJ C.L., DIERGAARDE P.J., HART M., PANNEKOEK H.;
 RL EMBO J. 5:3074-3074(1986).
 RN [5]
 RP SEQUENCE OF 764-2813.
 RM 86269895
 RA TITANI K., KUMAR S., TAKIO K., ERICSSON L.H., WADE R.D., ASHIDA K.,
 RA WALSH K.A., CHOPEK M.W., SADLER J.E., FUJIKAWA K.;
 RL BIOCHEMISTRY 25:3171-3184(1986).
 RN [6]
 RP SEQUENCE OF 781-1424 FROM N.A.
 RM 86269894
 RA SHELTON-INLOES B.B., TITANI K., SADLER J.E.;
 RL BIOCHEMISTRY 25:3164-3171(1986).
 RN [7]
 RP SEQUENCE OF 764-873 AND 1289-2813 FROM N.A.
 RM 86016708
 RA SADLER J.E., SHELTON-INLOES B.B., SORACE J.M., HARLAN J.M.,
 RA TITANI K., DAVIE E.W.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 82:6394-6398(1985).
 RN [8]
 RP SEQUENCE OF 2731-2813 FROM N.A.
 RM 85269603
 RA VERWEIJ C.L., DE VRIES C.J.M., DISTEL B., VAN ZONNEVELD A.-J.,
 RA VAN KESSEL A.G., VAN MOURIK J.A., PANNEKOEK H.;
 RL NUCLEIC ACIDS RES. 13:4699-4717(1985).
 RN [9]
 RP SEQUENCE OF 1-177 FROM N.A.
 RM 88111704
 RA BONTHON D., ORKIN S.H.;
 RL EUR. J. BIOCHEM. 171:51-57(1988).
 RN [10]
 RP STRUCTURE OF CARBOHYDRATES.
 RM 86274702
 RA SAMOR B., MICHALSKI J.C., DEBRAY H., MAZURIER C., GOUDEMAND M.,
 RA VAN HALBEK H., VLEGENTHART J.F.G., MONTREUIL J.;
 RL EUR. J. BIOCHEM. 158:295-298(1986).
 RN [11]
 RP VARIANT NORMANDY-1.

RM 91296824
RA TULEY E.A., GAUCHER C., JORIEUX S., MORRALL N.K., SADLER J.E.,
RA MAZURIER C.;
RL PROC. NATL. ACAD. SCI. U.S.A. 88:6377-6381(1991).
RN [12]
RP VARIANTS NORMANDY-2 AND NORMANDY-3.
RM 92001464
RA GAUCHER C., MERCIER B., JORIEUX S., OUFKIR D., MAZURIER C.;
RL BR. J. HAEMATOL. 78:506-514(1991).
RN [13]
RP VARIANT LEU-1266.
RM 93253064
RA HOLMBERG L., DENT J.A., SCHNEPPENHEIM R., BUDDÉ U., WARE J.,
RA RUGGERI Z.M.;
RL J. CLIN. INVEST. 91:2169-2177(1993).
RN [14]
RP VARIANT ASP-1268.
RM 93388627
RA RABINOWITZ I., RANDI A.M., SHINDLER K.S., TULEY E.A., RUSTAGI P.K.,
RA SADLER J.E.;
RL J. BIOL. CHEM. 268:20497-20501(1993).
RN [15]
RP VARIANT ARG-1272.
RM 93041232
RA LAVERGNE J.-M., DE PAILLETTE L., BAHNAK B.R., RIBBA A.-S.,
RA FRESSINAUD E., MEYER D., PIETU G.;
RL BR. J. HAEMATOL. 82:66-72(1992).
RN [16]
RP VARIANTS TRP-1306; CYS-1308 AND PRO-1613.
RM 91185601
RA RANDI A.M., RABINOWITZ I., MANCUSO D.J., MANNUCCI P.M., SADLER J.E.,
RL J. CLIN. INVEST. 87:1220-1226(1991).
RN [17]
RP VARIANTS TRP-1306; CYS-1308; LEU-1314 AND LEU-1318.
RM 93041230
RA DONNER M., KRISTOFFERSSON A.C., LENK H., SCHEIBEL E., DAHLBACK B.,
RA NILSSON I.M., HOLMBERG L.;
RL BR. J. HAEMATOL. 82:58-65(1992).
RN [18]
RP VARIANTS TRP-1306; MET-1316; THR-1628 AND SER-1648.
RM 93042596
RA PIETU G., RIBBA A.S., DE PAILLETTE L., CHEREL G., LAVERGNE J.M.,
RA BAHNAK B.R., MEYER D.;
RL BLOOD COAGUL. FIBRINOLYSIS 3:415-421(1992).
RN [19]
RP VARIANT CYS-1308.
RM 92104315
RA DONNER M., ANDERSSON A.-M., KRISTOFFERSSON A.-C., NILSSON I.M.,
RA DAHLBACK B., HOLMBERG L.;
RL EUR. J. HAEMATOL. 47:342-345(1991).
RN [20]
RP VARIANT CYS-1313.
RM 91187908
RA WARE J., DENT J.A., AZUMA H., SUGIMOTO M., KYRLE P.A., YOSHIOKA A.,
RA RUGGERI Z.M.;
RL PROC. NATL. ACAD. SCI. U.S.A. 88:2946-2950(1991).
RN [21]
RP VARIANT LEU-1314.
RM 93041230
RA DONNER M., KRISTOFFERSSON A.C., LENK H., SCHEIBEL E., DAHLBACK B.,

RA NILSSON I.M., HOLMBERG L.;
RL BR. J. HAEMATOL. 82:58-65(1992).
RN [22]
RP VARIANT MET-1316.
RM 92109240
RA MURRAY E.W., GILES A.R., LILLICRAP D.;
RL AM. J. HUM. GENET. 50:199-207(1992).
RN [23]
RP VARIANT SER-1324.
RM 93028560
RA RABINOWITZ I., TULEY E.A., MANCUSO D.J., RANDI A.M., FIRKIN B.G.,
RA HOWARD M.A., SADLER J.E.;
RL PROC. NATL. ACAD. SCI. U.S.A. 89:9846-9849(1992).
RN [24]
RP VARIANTS TRP-1306; CYS-1308; MET-1316; GLN-1341 AND HIS-1399.
RM 91185602
RA COONEY K.A., NICHOLS W.C., BRUCK M.E., BAHOU W.F., SHAPIRO A.D.,
RA BOWIE E.J.W., GRALNICK H.R., GINSBURG D.;
RL J. CLIN. INVEST. 87:1227-1233(1991).
RN [25]
RP VARIANT CYS-1514.
RM 93168632
RA GAUCHER C., HANSS M., DECHAVANNE M., MAZURIER C.;
RL BR. J. HAEMATOL. 83:94-99(1993).
RN [26]
RP VARIANTS TRP-1597 AND ASP-1607.
RM 89264495
RA GINSBURG D., KONKLE B.A., GILL J.C., MONTGOMERY R.R.,
RA BOCKENSTEDT P.L., JOHNSON T.A., YANG A.Y.;
RL PROC. NATL. ACAD. SCI. U.S.A. 86:3723-3727(1989).
RN [27]
RP VARIANTS GLN-1597; ARG-1609 AND GLU-1665.
RM 93333089
RA INBAL A., ENGLENDER T., KORNDROT N., RANDI A.M., CASTAMAN G.,
RA MANNUCCI P.M., SADLER J.E.;
RL BLOOD 82:830-836(1993).
RN [28]
RP VARIANT THR-1628.
RM 91196734
RA IANNUZZI M.C., HIDAKA N., BOEHKE M., BRUCK M.E., HANNA W.T.,
RA COLLINS F.S., GINSBURG D.;
RL AM. J. HUM. GENET. 48:757-763(1991).
RN [29]
RP VARIANT LYS-1638.
RM 93054652
RA RIBBA A.S., VOORBERG J., MEYER D., PANNEKOEK H., PIETU G.;
RL J. BIOL. CHEM. 267:23209-23215(1992).
RN [30]
RP VARIANT TYR-2739.
RM 94375013
RA ZHANG Z.P., BLOMBAECK M., EGBERG N., FALK G., ANVRET M.;
RL GENOMICS 21:188-193(1994).
CC -!- FUNCTION: VWF HAS AN IMPORTANT FUNCTION IN THE MAINTENANCE OF
CC HOMEOSTASIS, IT PARTICIPATES IN PLATELET-VESSEL WALL INTERACTIONS
CC BY FORMING A NONCOVALENT COMPLEX WITH COAGULATION FACTOR VIII AT
CC THE SITE OF VASCULAR INJURY.
CC -!- SUBUNIT: MULTIMERIC.
CC -!- TISSUE SPECIFICITY: BLOOD.
CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS.

CC -!- THE A DOMAINS APPEAR TO BE RELATED TO 225 RESIDUES OF COMPLEMENT
CC FACTOR B (THE LAST 5 RESIDUES OF BA AND THE AMINO-TERMINAL 220
CC RESIDUES OF BB).
CC -!- DISEASE: DEFICIENCIES IN VWF ARE ASSOCIATED WITH VARIOUS FORMS
CC OF VON WILLEBRAND DISEASE (VWD) CHARACTERIZED BY FREQUENT BLEEDING
CC (GINGIVAL, MINOR SKIN LACERATIONS, MENORRHAGIA, ETC.). TYPE I VWD
CC IS ASSOCIATED WITH A QUANTITATIVE DEFICIENCY OF VWF; TYPE II BY
CC NORMAL TO DECREASED PLASMA LEVEL OF VWF; TYPE III BY A VIRTUAL
CC ABSENCE OF VWF. THERE ARE SUBTYPES (A TO H) OF TYPE II VWD; FOR
CC EXAMPLE: TYPE IIA IS CHARACTERIZED BY THE ABSENCE OF VWF HIGH
CC MOLECULAR WEIGHT MULTIMERS IN PLASMA.
DR EMBL; M25828; HSVWFA01.
DR EMBL; M25829; HSVWFA02.
DR EMBL; M25830; HSVWFA03.
DR EMBL; M25831; HSVWFA04.
DR EMBL; M25832; HSVWFA05.
DR EMBL; M25833; HSVWFA06.
DR EMBL; M25834; HSVWFA07.
DR EMBL; M25835; HSVWFA08.
DR EMBL; M25836; HSVWFA09.
DR EMBL; M25837; HSVWFA10.
DR EMBL; M25838; HSVWFA11.
DR EMBL; M25839; HSVWFA12.
DR EMBL; M25840; HSVWFA13.
DR EMBL; M25841; HSVWFA14.
DR EMBL; M25842; HSVWFA15.
DR EMBL; M25843; HSVWFA16.
DR EMBL; M25844; HSVWFA17.
DR EMBL; M25845; HSVWFA18.
DR EMBL; M25846; HSVWFA19.
DR EMBL; M25847; HSVWFA20.
DR EMBL; M25848; HSVWFA21.
DR EMBL; M25849; HSVWFA22.
DR EMBL; M25850; HSVWFA23.
DR EMBL; M25851; HSVWFA24.
DR EMBL; M25852; HSVWFA25.
DR EMBL; M25853; HSVWFA26.
DR EMBL; M25854; HSVWFA27.
DR EMBL; M25855; HSVWFA28.
DR EMBL; M25856; HSVWFA29.
DR EMBL; M25857; HSVWFA30.
DR EMBL; M25858; HSVWFA31.
DR EMBL; M25859; HSVWFA32.
DR EMBL; M25860; HSVWFA33.
DR EMBL; M25861; HSVWFA34.
DR EMBL; M25862; HSVWFA35.
DR EMBL; M25863; HSVWFA36.
DR EMBL; M25864; HSVWFA37.
DR EMBL; M25865; HSVWFA38.
DR EMBL; M10321; HSVWF1.
DR EMBL; M10320; HSVWF2.
DR EMBL; X04146; HSVWFR.
DR EMBL; X02672; HSVWFC.
DR EMBL; X04385; HSVWFR1.
DR EMBL; X06828; HSVWF123.
DR EMBL; X06829; HSVWF45.
DR PIR; A03212; VWHU.
DR PIR; A34480; A34480.
DR PIR; A25298; A25298.
DR PIR; A25469; A25469.

DR PIR; S02377; S02377.
DR PIR; S07363; S07363.
DR MIM; 193400; 11TH EDITION.
DR MIM; 277480; 11TH EDITION.
KW BLOOD COAGULATION; PLATELET; GLYCOPROTEIN; EXTRACELLULAR MATRIX;
KW PLASMA; ENDOTHELIAL CELL; REPEAT; CELL ADHESION; SIGNAL;
KW DISEASE MUTATION; POLYMORPHISM; VON WILLEBRAND DISEASE.
FT SIGNAL 1 22
FT PROPEP 23 763 D1/D2.
FT CHAIN 764 2813 VON-WILLEBRAND FACTOR.
FT DOMAIN 764 787 AMINO-TERMINAL.
FT DOMAIN 788 833 E1.
FT DOMAIN 826 853 CX.
FT DOMAIN 842 1130 D3.
FT DOMAIN 1260 1479 A1.
FT DOMAIN 1480 1672 A2.
FT DOMAIN 1673 1874 A3.
FT DOMAIN 1934 2203 D4.
FT DOMAIN 2216 2261 E2.
FT DOMAIN 2296 2330 B1.
FT DOMAIN 2340 2365 B2.
FT DOMAIN 2375 2399 B3.
FT DOMAIN 2400 2515 C1.
FT DOMAIN 2544 2662 C2.
FT DOMAIN 2663 2813 CARBOXYL-TERMINAL.
FT REPEAT 2216 2261 E DOMAIN.
FT REPEAT 826 853 E DOMAIN.
FT REPEAT 2400 2515 C DOMAIN.
FT REPEAT 2544 2662 C DOMAIN.
FT REPEAT 842 1130 D DOMAIN.
FT REPEAT 1934 2203 D DOMAIN.
FT REPEAT 1260 1479 A DOMAIN.
FT REPEAT 1480 1672 A DOMAIN.
FT REPEAT 1673 1874 A DOMAIN.
FT REPEAT 2296 2330 B DOMAIN.
FT REPEAT 2340 2365 B DOMAIN.
FT REPEAT 2375 2399 B DOMAIN.
FT SIMILAR 1260 1479 WITH COMPLEMENT FACTOR B.
FT SIMILAR 1480 1672 WITH COMPLEMENT FACTOR B.
FT SIMILAR 1673 1874 WITH COMPLEMENT FACTOR B.
FT SITE 2507 2509 CELL ATTACHMENT SITE.
FT DISULFID 1272 1458
FT CARBOHYD 857 857
FT CARBOHYD 1147 1147
FT CARBOHYD 1231 1231
FT CARBOHYD 1248 1248
FT CARBOHYD 1255 1255
FT CARBOHYD 1256 1256
FT CARBOHYD 1263 1263
FT CARBOHYD 1468 1468
FT CARBOHYD 1477 1477
FT CARBOHYD 1486 1486
FT CARBOHYD 1487 1487
FT CARBOHYD 1515 1515
FT CARBOHYD 1574 1574
FT CARBOHYD 1679 1679
FT CARBOHYD 2223 2223
FT CARBOHYD 2290 2290
FT CARBOHYD 2298 2298
FT CARBOHYD 2298 2298

FT CARBOHYD 2357 2357
 FT CARBOHYD 2400 2400
 FT CARBOHYD 2546 2546
 FT CARBOHYD 2585 2585
 FT CARBOHYD 2790 2790
 FT CARBOHYD 789 789
 FT VARIANT 791 791
 FT VARIANT 816 816
 FT VARIANT 852 852
 FT VARIANT 854 854
 FT VARIANT 857 857
 FT VARIANT 1266 1266
 FT VARIANT 1268 1268
 FT VARIANT 1272 1272
 FT VARIANT 1306 1306
 FT VARIANT 1308 1308
 FT VARIANT 1313 1313
 FT VARIANT 1314 1314
 FT VARIANT 1316 1316
 FT VARIANT 1318 1318
 FT VARIANT 1324 1324
 FT VARIANT 1341 1341
 FT VARIANT 1381 1381
 FT VARIANT 1399 1399
 FT VARIANT 1514 1514
 FT VARIANT 1597 1597
 FT VARIANT 1597 1597
 FT VARIANT 1607 1607
 FT VARIANT 1609 1609
 FT VARIANT 1633 1633
 FT VARIANT 1638 1638
 FT VARIANT 1648 1648
 FT VARIANT 1665 1665
 FT VARIANT 1665 1665
 FT VARIANT 2739 2739
 FT CONFLICT 471 471
 FT CONFLICT 484 484
 FT CONFLICT 770 770
 FT CONFLICT 804 804
 FT CONFLICT 1472 1472
 FT CONFLICT 1914 1914
 FT CONFLICT 2168 2168
 SQ SEQUENCE 2813 AA; 309296 MW; 19889791 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
 Residue Identity = 42% Matches = 3 Mismatches = 4
 Gaps = 0 Conservative Substitutions = 0

X X
 CGTXXXX
 |||
 NCPKGVYLCQGTFCNLTCSLSYPDE
 660 670

11. US-08-121-713B-3 (1-7)
 IP3R_MOUSE INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN TYPE
 ID IP3R_MOUSE STANDARD; PRT; 2749 AA.
 AC F11881;

DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN TYPE 1 RECEPTOR
 DE (INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN P(400)) (TYPE 1 INSP3
 DE RECEPTOR).
 CN INSP3R.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=CEREBELLA PURKINJE NEURONS;
 RM 90044039
 RA FURUICHI T., YOSHIKAWA S., MIYAWAKI A., WADA K., MAEDA N.,
 RA MIKOSHIBA K.;
 RL NATURE 342:32-38(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=CEREBELLUM;
 RM 89345101
 RA FURUICHI T., YOSHIKAWA S., MIKOSHIBA K.;
 RL NUCLEIC ACIDS RES. 17:5385-5386(1989).
 RN [3]
 RP ALTERNATIVE SPLICING.
 RM 91296797
 RA NAKAGAWA T., OKANO H., FURUICHI T., ARUGA J., MIKOSHIBA K.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:6244-6248(1991).
 CC -1- FUNCTION: RECEPTOR FOR INOSITOL 1,4,5-TRISPHOSPHATE, A SECOND
 CC MESSENGER THAT MEDIATES THE RELEASE OF INTRACELLULAR CALCIUM.
 CC THE RECEPTOR CONTAINS A CALCIUM CHANNEL IN ITS C-TERMINAL
 CC EXTREMITY. ITS LARGE N-TERMINAL CYTOPLASMIC REGION HAS THE LIGAND-
 CC BINDING SITE IN THE N-TERMINUS AND MODULATORY SITES IN THE MIDDLE
 CC PORTION IMMEDIATELY UPSTREAM OF THE CHANNEL REGION.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- PTM: PHOSPHORYLATED BY CYCLIC-AMP KINASE. PHOSPHORYLATION PREVENTS
 CC THE LIGAND-INDUCED OPENING OF THE CALCIUM CHANNELS.
 CC -1- CALCIUM APPEARS TO INHIBIT LIGAND BINDING TO THE RECEPTOR, MOST
 CC PROBABLY BY INTERACTING WITH A DISTINCT CALCIUM-BINDING PROTEIN
 CC WHICH THEN INHIBITS THE RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- ALTERNATIVE PRODUCTS: ADDITIONAL SUBTYPES OF INSP3R ARISE BY
 CC ALTERNATIVE SPLICING OF THE SAME GENE.
 CC -1- SIMILARITY: TO RYANODINE RECEPTOR.
 DR EMBL; X15373; MWP400.
 DR PIR; S04844; ACMSIT.
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION;
 KW ENDOPLASMIC RETICULUM; IONIC CHANNEL; ION TRANSPORT; CALCIUM CHANNEL;
 KW ALTERNATIVE SPLICING.
 FT DOMAIN 1 2273 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 2274 2294 M1 (POTENTIAL).
 FT TRANSMEM 2308 2326 M2 (POTENTIAL).
 FT TRANSMEM 2334 2356 M3 (POTENTIAL).
 FT TRANSMEM 2365 2387 M4 (POTENTIAL).
 FT TRANSMEM 2391 2407 M5 (POTENTIAL).
 FT TRANSMEM 2440 2462 M6 (POTENTIAL).
 FT TRANSMEM 2530 2549 M7 (POTENTIAL).
 FT TRANSMEM 2570 2589 M8 (POTENTIAL).
 FT DOMAIN 2590 2749 CYTOPLASMIC (POTENTIAL).

FT VARSPLIC 318 332 MISSING (IN INSP3R SI-).
FT VARSPLIC 1692 1731 MISSING (IN INSP3R SIABC-).
FT VARSPLIC 1715 1715 MISSING (IN INSP3R SIIB-).
FT VARSPLIC 1715 1731 MISSING (IN INSP3R SIIBG-).
FT MOD_RES 1588 1588 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT MOD_RES 1755 1755 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
SQ SEQUENCE 2749 AA; 313193 MW; 22837568 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
|||

IIIFMSFVGNCGTFTTRGYRAMVLDFEF
2370 X 2380 2390

12. US-08-121-713B-3 (1-7)
IP3R_RAT INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN TYPE

ID IP3R_RAT STANDARD; PRT; 2749 AA.
AC R29994;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN TYPE 1 RECEPTOR
DE (INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN P(400)) (TYPE 1 INSP3
DE RECEPTOR).
GN INSP3R.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RM 90324264
RA MIGNERY G.A., NEWTON C.L., ARCHER B.T. III, SOEDHOF T.C.;
RL J. BIOL. CHEM. 265:12679-12685(1990).
CC -!- FUNCTION: RECEPTOR FOR INOSITOL 1,4,5-TRISPHOSPHATE, A SECOND
CC MESSENGER THAT MEDIATES THE RELEASE OF INTRACELLULAR CALCIUM.
CC THE RECEPTOR CONTAINS A CALCIUM CHANNEL IN ITS C-TERMINAL
CC EXTREMITY. ITS LARGE N-TERMINAL CYTOPLASMIC REGION HAS THE LIGAND-
CC BINDING SITE IN THE N-TERMINUS AND MODULATORY SITES IN THE MIDDLE
CC PORTION IMMEDIATELY UPSTREAM OF THE CHANNEL REGION.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- PTM: PHOSPHORYLATED BY CYCLIC-AMP KINASE. PHOSPHORYLATION PREVENTS
CC THE LIGAND-INDUCED OPENING OF THE CALCIUM CHANNELS.
CC -!- CALCIUM APPEARS TO INHIBIT LIGAND BINDING TO THE RECEPTOR, MOST
CC PROBABLY BY INTERACTING WITH A DISTINCT CALCIUM-BINDING PROTEIN
CC WHICH THEN INHIBITS THE RECEPTOR.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED.
CC -!- SIMILARITY: TO RYANODINE RECEPTOR.
DR EMBL; J05510; RRI145TR.
DR PIR; A36579; A36579.
KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION;
KW ENDOPLASMIC RETICULUM; IONIC CHANNEL; ION TRANSPORT; CALCIUM CHANNEL;
KW ALTERNATIVE SPLICING.

FT DOMAIN 1 2273 CYTOPLASMIC (POTENTIAL).
FT TRANSEM 2274 2294 M1 (POTENTIAL).
FT TRANSEM 2308 2326 M2 (POTENTIAL).
FT TRANSEM 2334 2356 M3 (POTENTIAL).
FT TRANSEM 2365 2387 M4 (POTENTIAL).
FT TRANSEM 2391 2407 M5 (POTENTIAL).
FT TRANSEM 2440 2462 M6 (POTENTIAL).
FT TRANSEM 2530 2549 M7 (POTENTIAL).
FT TRANSEM 2570 2589 M8 (POTENTIAL).
FT DOMAIN 2590 2749 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 1589 1589 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT MOD_RES 1755 1755 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT VARSPLIC 322 336 MISSING (IN VARIANT).
FT VARIANT 1372 1372 MISSING (IN ALL, BUT P17 CLONES).
SQ SEQUENCE 2749 AA; 313132 MW; 22831603 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
|||

IIIFMSFVGNCGTFTTRGYRAMVLDFEF
2370 X 2380 2390

13. US-08-121-713B-3 (1-7)
ZEP1_HUMAN ZINC FINGER PROTEIN 40 (HUMAN IMMUNODEFICIENCY VIR

ID ZEP1_HUMAN STANDARD; PRT; 2717 AA.
AC P15822;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ZINC FINGER PROTEIN 40 (HUMAN IMMUNODEFICIENCY VIRUS TYPE I ENHANCER-
DE BINDING PROTEIN 1) (HIV-EF1) (MAJOR HISTOCOMPATIBILITY COMPLEX BINDING
DE PROTEIN 1) (MBP-1) (POSITIVE REGULATORY DOMAIN II BINDING FACTOR 1)
DE (PDI1-BF1).
GN HIVEP1 OR ZNF40.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RM 90169514
RA FAN C.M., MANIATIS T.;
RL GENES DEV. 4:29-42(1990).
RN [2]
RP STRUCTURE BY NMR OF 2113-2142.
RM 91064333
RA OMICHINSKI J.G., CLORE G.M., APPELLA E., SAKAGUCHI K.,
RA GRONENBORN A.M.;
RL BIOCHEMISTRY 29:9324-9334(1990).
RN [3]
RP STRUCTURE BY NMR OF 2087-2142.
RM 92232684
RA OMICHINSKI J.G., CLORE G.M., ROBIEN M., SAKAGUCHI K., APPELLA E.,
RA GRONENBORN A.M.;
RL BIOCHEMISTRY 31:3907-3917(1992).

CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
CC 5'-GGGACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
CC MHC. INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
CC IN T-CELL ACTIVATION.
CC
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY SEPARATED
CC AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH ZINC-FINGER
CC IN-BETWEEN.
CC -1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
CC -1- SIMILARITY: 70% BETWEEN THE TWO ZINC-FINGER DOMAINS.
CC EMBL; X51435; HSZFPBF1.
CC DR PIR; A34203; A34203.
CC DR PDB; 4ZNF; 15-JAN-92.
CC DR PDB; 4ZNF; 15-JAN-92.
CC DR PDB; 1BBO; 31-OCT-93.
CC DR TRANSFAC; T00497; -.
CC DR MIM; 194540; 11TH EDITION.
CC DR PROSITE; PS00028; ZINC FINGER C2H2.
CC KW TRANSCRIPTION REGULATION; ZINC-FINGER; METAL-BINDING; DNA-BINDING;
CC KW NUCLEAR PROTEIN; 3D-STRUCTURE.
CC FT DOMAIN 406 436
CC FT ZN FING 406 428
CC FT ZN FING 434 456
CC FT ZN FING 955 981
CC FT DOMAIN 2087 2139
CC FT ZN FING 2087 2109
CC FT ZN FING 2115 2139
CC FT STRAND 2088 2088
CC FT TURN 2090 2092
CC FT STRAND 2095 2095
CC FT HELIX 2099 2108
CC FT TURN 2109 2109
CC FT STRAND 2115 2116
CC FT STRAND 2123 2124
CC FT HELIX 2127 2135
CC SEQUENCE 2717 AA; 297217 MW; 19898490 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
Residue Identity = 42% Matches = 3 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
CGTXXX
|||
TDNSECISHC GTTSPSYNTAFDVL
180 X 190 200

14. US-08-121-713B-3 (1-7)
G156_PARP 156G SURFACE PROTEIN PRECURSOR.

ID G156_PARP STANDARD; PRT; 2715 AA.
AC P13837;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-AUG-1990 (REL. 15, LAST ANNOTATION UPDATE)

DE 156G SURFACE PROTEIN PRECURSOR.
GN 156G.
OS PARAMECIUM PRIMAURELIA.
OC EUKARYOTA; PROTOZOA; CILIOPHORA; CILIATA; HOLOTRICHA; HYMENOSTOMATIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=156;
RM 87060934
RA PRAT A., KATINKA M., CARON F., MEYER E.;
RL J. MOL. BIOL. 189:47-60(1986).
CC -1- FUNCTION: THIS PROTEIN IS THE SURFACE ANTIGEN OR IMMOBILIZATION
CC ANTIGEN OF PARAMECIUM PRIMAURELIA.
CC -1- IT HAS INTERNAL HOMOLOGIES AND A HIGHLY PERIODIC STRUCTURE WITH
CC 37 PERIODS OF ABOUT 75 RESIDUES, EACH PERIOD CONTAINING 8
CC CYSTEINES, EXCEPT FOR FOUR HALF PERIODS. A VARIABLE PART OF 475
CC RESIDUES COMPRISES 4 ALMOST IDENTICAL PERIODS IN THE MIDDLE OF THE
CC PROTEIN.
CC -1- EXPRESSION OF G PROTEIN OCCURS AT LOW TEMPERATURES (14-32
CC DEGREE CELSIUS).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- SIMILARITY: 98% TO THE ALLELIC FORM 168G PROTEIN (P17053) IN
CC PERIODIC STRUCTURE AND 80% IN VARIABLE DOMAIN IN THE MIDDLE OF
CC THE PROTEIN.
CC EMBL; X03882; PPSGP.
CC DR PIR; A23475; A23475.
CC DR HSSP; P06620; 1INA.
CC KW SIGNAL; REPEAT; ANTIGEN; MEMBRANE; GPI-ANCHOR.
CC FT SIGNAL 1 20 POTENTIAL.
CC FT CHAIN 21 2715 156G SURFACE PROTEIN.
CC FT DOMAIN 106 2560 37 X 75 AA APPROXIMATE REPEATS.
CC FT SIMILAR 1 222 88% TO PARAMECIUM TETRAURELIA A
CC FT PROTEIN.
CC SEQUENCE 2715 AA; 279551 MW; 16284274 CN;

X X
CGTXXX
|||

ITGTGLTTAICGTDAGCVANVNGTAC
1070 1080 X 1090

15. US-08-121-713B-3 (1-7)
CENP_HUMAN CENTROMERIC PROTEIN E (CENP-E PROTEIN).

ID CENP_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE DE CENTROMERIC PROTEIN E (CENP-E PROTEIN).
GN CENPE.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.

RM 93024922
 RA YEN T.J., LI G., SCHAAR B.T., SZILAK I., CLEVELAND D.W.;
 RL NATURE 359:536-539(1992).
 CC -!- FUNCTION: PROBABLE KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE
 CC MITOSIS AT THE G2 PHASE OF THE CELL CYCLE. PROBABLY IMPORTANT
 CC FOR CHROMOSOME MOVEMENT AND/OR SPINDLE ELONGATION.
 CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
 CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -!- SIMILARITY: BELONGS TO A FAMILY OF KINESIN-LIKE PROTEINS
 CC CHARACTERIZED BY THE PRESENCE OF A MECHANOCHEMICAL DOMAIN TETHERED
 CC TO DIFFERENT PROTEIN BINDING DOMAINS.
 CC EMBL; Z15005; HSCENPE.
 DR PIR; S28261; S28261.
 DR MIM; 117143; 11TH EDITION.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN.
 KW MOTOR PROTEIN; CELL DIVISION; ATP-BINDING; COILED COIL; MITOSIS;
 KW CELL CYCLE; CENTROMERE.
 FT DOMAIN 1 335 MECHANOCHEMICAL (MOTOR).
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
 FT NP BIND 86 93 ATP (BY SIMILARITY).
 SQ SEQUENCE 2663 AA; 312087 MW; 23820152 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.98
 Residue Identity = 42% Matches = 3 Mismatches = 4
 Gaps = 0 Conservative Substitutions = 0

X X
 CGTXXXX
 |||
 IYNETITDLICGTQKPKLIIRDVNR
 140 X 150 160

maryh@stic

stdin

NeWSprinter20

Fri May 19 10:59:04 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5

9. R55273 Beta subunit of integrin cell 1822 7 7 2.92 0
 10. R59751 Type II collagen. 1418 7 7 2.92 0
 11. R60292 Varicella zoster virus IEP175 1310 7 7 2.92 0
 12. R54841 HER4. 1308 7 7 2.92 0
 13. R52701 Plasmid pASK60-Strep reading 1277 7 7 2.92 0
 14. R52702 Plasmid pASK60-Strep reading 1277 7 7 2.92 0
 15. R54074 CryET5. 1229 7 7 2.92 0
 16. R55764 Sequence encoded by the cDNA 1153 7 7 2.92 0
 17. R55272 Alpha subunit of integrin cel 1073 7 7 2.92 0
 18. R54843 HER4 with alternate 3'-end wi 1058 7 7 2.92 0
 19. R54066 Non-A, non-B hepatitis virus 1051 7 7 2.92 0
 20. R54067 Non-A, non-B hepatitis virus 1031 7 7 2.92 0
 21. R53921 HCV fusion protein corresp. t 980 7 7 2.92 0
 22. R52566 Phosphoenol pyruvate carboxyl 964 7 7 2.92 0
 23. R52567 Phosphoenol pyruvate carboxyl 964 7 7 2.92 0
 24. R56990 Bacillus deramificans pullula 957 7 7 2.92 0
 25. R56989 Bacillus deramificans mature 928 7 7 2.92 0
 26. R53963 VEF. 901 7 7 2.92 0
 27. R60252 NF-ATp. 890 7 7 2.92 0
 28. R59924 Human GAP protein. 870 7 7 2.92 0
 29. R53732 S. cerevisiae Plci protein. 869 7 7 2.92 0
 30. R53404 S-Locus receptor (serine/thre 858 7 7 2.92 0
 31. R53228 Rice starch branching enzyme. 820 7 7 2.92 0
 32. R55692 hnRNP U protein. 806 7 7 2.92 0
 33. R52783 Bovine enterokinase. 798 7 7 2.92 0
 34. R52690 HCV CKS-33-BCD antigen. 781 7 7 2.92 0
 35. R55708 Glycosyltransferase hybrid. 767 7 7 2.92 0
 36. R55709 Glycosyltransferase hybrid. 767 7 7 2.92 0
 37. R54611 Native CD26. 766 7 7 2.92 0
 38. R53778 Sequence of human amyloid pro 763 7 7 2.92 0
 39. R54612 Delta3-9 CD26. 759 7 7 2.92 0
 40. R59784 Human pro-hormone convertase 753 7 7 2.92 0
 41. R67764 Murine pro-hormone convertase 753 7 7 2.92 0
 42. R55097 Prostate-specific membrane an 750 7 7 2.92 0
 43. R55206 Human zona pellucida ZPA prot 745 7 7 2.92 0
 44. R54867 Hepatitis C Virus core protei 743 7 7 2.92 0
 45. R54613 Delta24-34 CD26. 739 7 7 2.92 0

1. US-08-121-713B-7 (1-7)
 R54099 NANBHV El/E2 protein.
 ID R54099 standard; Protein; 3014 AA.
 AC R54099;
 DT 09-FEB-1995 (first entry)
 DE NANBHV El/E2 protein.
 KW El/E2 protein; non-A, non-B hepatitis virus; NANBHV; signal peptide;
 KW heterogenic; virus; transformation; insect cell; antigen; anti-NANBHV;
 KW antibody; hepatitis C virus; HCV; vaccine.
 OS Non-A, non-B hepatitis virus.
 FH Key location/Qualifiers
 FT Peptide 832..847
 FT /note= "Peptide fragment not given in the specification,
 FT but encoded by the given cDNA sequence"
 FT Peptide 1296..1311
 FT /note= "Peptide fragment not given in the specification,
 FT but encoded by the given cDNA sequence"
 FT Peptide 1760..1775
 FT /note= "Peptide fragment not given in the specification,
 FT but encoded by the given cDNA sequence"

FT Peptide 2688..2703
 FT /note= "Peptide fragment not given in the specification,
 FT but encoded by the given cDNA sequence"
 PN J06141873-A.
 PD 24-MAY-1994.
 PF 13-MAR-1992; 089371.
 PR 13-MAR-1992; JP-089371.
 PA (KAGA) 2H KAGAKU & KESSEI RYOHO KENKYUSHO.
 DR WPI; 94-205030/25.
 DR N-PSDB; 064175.
 PT Virus vector contg hepatitis C virus and signal sequence - useful
 PT in vaccines and in immunological detection
 PS Disclosure; Page 7-19; 23pp; Japanese.
 CC This sequence represents the El/E2 protein from non-A, non-B hepatitis
 CC virus (NANBHV). The cDNA encoding this sequence may be linked to a
 CC signal peptide (r54100) from a heterogenic virus so the the NANBHV
 CC protein may be expressed by a transformed insect cell. This protein
 CC may be used as an antigen in the generation of anti-NANBHV antibodies.
 CC These antibodies may be used in a hepatitis C vaccine.
 SQ Sequence 3014 AA;
 SQ 293A; 178R; 91 N; 127D; 0 B; 97 C; 88 Q; 112E; 0 Z; 252G; 68 H;
 SQ 129I; 301L; 95 K; 61 M; 86 F; 205P; 216S; 216T; 65 W; 99 Y; 235V;
 Initial Score = 7 Optimized Score = 7 Significance = 2.92
 Residue Identity = 14% Matches = 1 Mismatches = 6
 Gaps = 0 Conservative Substitutions = 0
 X X X
 XXXXPXX
 RVLEDGVNATCNMFGCSFIFLLALL
 160 X 170 X 180
 2. US-08-121-713B-7 (1-7)
 R53417 Blood transmissible NANBHV protein.
 ID R53417 standard; Protein; 3010 AA.
 AC R53417;
 DT 17-JAN-1995 (first entry)
 DE Blood transmissible NANBHV protein.
 KW Polymerase chain reaction; PCR; amplify; primer; non-A, non-B hepatitis;
 KW NANBH; virus; blood transmissible; detection; hepatitis virus; RT-PCR;
 KW C100 antibody; HCV RNA; NS5 region.
 OS Non-A, non-B hepatitis virus.
 FH Key location/Qualifiers
 FT Misc_difference 222
 FT /label= His, Arg
 FT Misc_difference 226
 FT /label= Cys, Arg
 FT Misc_difference 246
 FT /label= Leu, Phe
 FT Misc_difference 263
 FT /label= Asp, Asn
 FT Misc_difference 291
 FT /label= Phe, Ser
 FT Misc_difference 311
 FT /label= Gly, Asp
 FT Misc_difference 398
 FT /label= Ser, Arg, Gly

FT Misc difference 400
FT /label= Thr, Ala
FT Misc difference 405
FT /label= Gln, Pro, Leu
FT Misc difference 410
FT /label= Lys, Arg
FT Misc difference 418
FT /label= Gly, Asp
FT Misc difference 430
FT /label= Asn, Asp
FT Misc difference 438
FT /label= Phe, Leu
FT Misc difference 478
FT /label= Arg, Lys
FT Misc difference 759
FT /label= Leu, Val
FT Misc difference 1017
FT /label= Ser, Asn
FT Misc difference 1036
FT /label= Thr, Ala
FT Misc difference 1056
FT /label= Glu, Asp
FT Misc difference 1201
FT /label= Met, Thr
FT Misc difference 1205
FT /label= Met, Ile
FT Misc difference 1255
FT /label= Asn, Tyr
FT Misc difference 1263
FT /label= Gly, Asp
FT Misc difference 1455
FT /label= Asn, Asp
FT Misc difference 1828
FT /label= Ala, Thr
FT Misc difference 1895
FT /label= Gly, Arg
FT Misc difference 1896
FT /label= Gly, Ile
FT Misc difference 2143
FT /label= Glu, Val
FT Misc difference 2144
FT /label= Asp, Glu
FT Misc difference 2462
FT /label= Cys, Arg
FT Misc difference 2486
FT /label= Val, Met
FT Misc difference 2488
FT /label= Lys, Gln
FT Misc difference 2844
FT /label= Leu, Met
FT Misc difference 2862
FT /label= Leu, Gln
FT Misc difference 2917
FT /label= Arg, Leu
FT Misc difference 2968
FT /label= Ser, Gly
FT Misc difference 2989
FT /label= Cys, Arg
FT Misc difference 2990
FT /label= Tyr, Cys

PN J06105690-A.
PD 19-APR-1994.
PF 10-MAR-1992; 051885.
PR 10-MAR-1992; JP-051885.
PA (KAEN/) KAENNO K.
DR WPI; 94-163130/20.
DR N-FSDB; Q63499.
PT Blood-transmissible non-A non-B hepatitis virus DNA - used for
PT detection of hepatitis virus
PS Claim 1; Page 8-20; 22pp; Japanese.
CC This sequence is encoded by the genome of a blood transmissible non-A,
CC non-B hepatitis (NANBH) virus. The cDNA sequence was isolated using the
CC primers given in Q63500-35. The amplified fragments are used in the
CC detection of hepatitis virus. The target DNA was isolated from serum
CC of chronically infected NANBH patients who were C100 antibody-positive
CC and HCV RNA (NS5 region) positive. Reverse transcription-PCR and PCR
CC were performed on cDNA and the total human NANBH DNA was constructed
CC from 23 clones.
SQ Sequence 3010 AA;
SQ 289A; 166R; 80 N; 124D; 0 B; 96 C; 94 Q; 113E; 0 Z; 241G; 63 H;
SQ 134I; 295L; 97 K; 57 M; 88 F; 204P; 216S; 225T; 64 W; 100Y; 227V;
SQ 37 Others;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPX
|
RKHPATYTKCGSGFWLTPRCIVDYPY
590 600 X 610

3. US-08-121-713B-7 (1-7)
R59921 RAS associated GAP NF201.

ID R59921 standard; protein; 2485 AA.
AC R59921;
DT 22-FEB-1995 (first entry)
DE RAS associated GAP NF201.
KW RAS; GTPase activating protein; GAP; GAP related domain; GRD;
KW pK10; pK11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;
KW neurofibromatosis type 1; NF1.
OS Homo sapiens.
PN WO9416069-A.
PD 21-JUL-1994.
PR 12-JAN-1994; U00198.
PR 15-JAN-1993; US-004824.
PA (SCHE) SCHERING CORP.
PI Kaziro Y, Nakafuku M;
DR WPI; 94-249216/30.
PT Blocking Ras-induced effects on a cell - by introducing a GTPase
PT activating protein to the cell, used esp. in treatment of cancers
PS Disclosure; Page 36-44; 87pp; English.
CC Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD) was
CC cloned into the yeast expression vector pK10 to obtain pK11. The
CC pK11 DNA was mutagenized by hydroxylamine in vitro and transformed
CC into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type
CC RAS2Val19 mutation. The heat shock sensitivity of the clones was

CC checked. Plasmid DNAs were recovered, re-transformed into TK161-
 CC R2V-D, and phenotypic reversion was examined. 2 Clones, NF201 and
 CC NF204 (given in R59922), which had strong suppression activity for
 CC RAS2Val19, were selected. The mutant NF1-GRDs were also able to
 CC inhibit v-Ras-induced transformation in mammalian cells.
 SQ Sequence 2485 AA;
 SQ 159A; 109R; 110N; 113D; 0 B; 51 C; 111Q; 146E; 0 Z; 110G; 80 H;
 SQ 127I; 312L; 141K; 75 M; 104F; 112P; 221S; 156T; 28 W; 63 Y; 157V;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
 Residue Identity = 14% Matches = 1 Mismatches = 6
 Gaps = 0 Conservative Substitutions = 0

LITGLVQLVQSHMPEIAQAEAMEALLV
 200 210

4. US-08-121-713B-7 (1-7)
 R59922 RAS associated GAP NF204.

ID R59922 standard; protein; 2485 AA.
 AC R59922;
 DT 22-FEB-1995 (first entry)
 DE RAS associated GAP NF204.
 KW Ras; GTPase activating protein; GAP; GAP related domain; GRD;
 KW pk110; pk11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;
 KW neurofibromatosis type 1; NF1.
 OS Homo sapiens.
 PN WO9416069-A.
 PD 21-JUL-1994.
 PF 12-JAN-1994; 000198.
 PR 15-JAN-1993; US-004824.
 PA (SCHE) SCHERING CORP.
 PI Kaziro Y, Nakafuku M;
 DR WPI; 94-249216/30.
 PT Blocking Ras-induced effects on a cell - by introducing a GTPase
 PT activating protein to the cell, used esp. in treatment of cancers
 PS Disclosure; Page 44-52; 87pp; English.

CC Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD) was
 CC cloned into the yeast expression vector pk110 to obtain pk11. The
 CC pk11 DNA was mutagenized by hydroxylamine in vitro and transformed
 CC into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type
 CC RAS2Val19 mutation. The heat shock sensitivity of the clones was
 CC checked. Plasmid DNAs were recovered, re-transformed into TK161-
 CC R2V-D, and phenotypic reversion was examined. 2 Clones, NF201
 CC (given in R59921) and NF204, which had strong suppression activity
 CC for RAS2Val19, were selected. The mutant NF1-GRDs were also able
 CC to inhibit v-Ras-induced transformation in mammalian cells.
 SQ Sequence 2485 AA;
 SQ 160A; 108R; 110N; 113D; 0 B; 52 C; 111Q; 146E; 0 Z; 109G; 81 H;
 SQ 125I; 313L; 136K; 76 M; 106F; 114P; 221S; 156T; 28 W; 63 Y; 157V;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
 Residue Identity = 14% Matches = 1 Mismatches = 6
 Gaps = 0 Conservative Substitutions = 0

X X X

XXXXPXX
 |
 LITGLVQLVQSHMPEIAQAEAMEALLV
 200 210

5. US-08-121-713B-7 (1-7)
 R55694 Carbamoyl-phosphate-synthetase II.

ID R55694 standard; Protein; 2391 AA.
 AC R55694;
 DT 06-DEC-1994 (first entry)
 DE Carbamoyl-phosphate-synthetase II.
 KW Carbamoyl-phosphate-synthetase II; CPSII; psCPSII gene;
 KW malaria.
 OS Plasmodium falciparum.

FH Key Location/Qualifiers
 FT Domain 1..690
 FT /note= "glutamine-amidotransferase domain"
 FT Domain 1..270
 FT /note= "structural subdomain"
 FT Peptide 271..482
 FT /note= "insert sequence"
 FT Domain 483..690
 FT /note= "glutaminase subdomain"
 FT Domain 691..2391
 FT /note= "carbamoyl-phosphate-synthase domain"
 FT Domain 691..1254
 FT /note= "ATP binding subdomain CPSa"
 FT Peptide 1255..1857
 FT /note= "insert sequence"
 FT Domain 1858..2391
 FT /note= "ATP binding subdomain CPSb"
 PN WO9412643-A.

PD 09-JUN-1994.
 PF 02-DEC-1993; AU0617.
 PR 03-DEC-1992; AU-006206.
 PR 16-DEC-1992; AU-006380.
 PA (UNIX) UNISEARCH LTD.
 PI Flores MV, Osullivan WJ, Stewart TS;
 DR WPI; 94-200271/24.
 DR N-PSDB; Q62924.
 PT Nucleic acid encoding carbamoyl phosphate synthetase II -
 PT isolated from Plasmodium falciparum, used to develop prods. for
 PT the treatment of malaria.
 PS Disclosure; Page 6-16; 31pp; English.

CC The cDNA sequence encoding the carbamoyl-phosphate-transferase II
 CC (CPSII) of Plasmodium falciparum was determined. The cDNA encodes
 CC a protein that includes 2 insert sequences not found in other CPSII
 CC proteins. The first separates the putative structural subdomain and
 CC the glutaminase subdomain of the glutamine-amidotransferase subunit
 CC of CPSII, while the second separates 2 ATP binding subdomains of the
 CC CPSII subunit, CPSa and CPSb.
 SQ Sequence 2391 AA;
 SQ 68 A; 61 R; 328N; 150D; 0 B; 53 C; 48 Q; 164E; 0 Z; 110G; 53 H;
 SQ 219I; 185L; 234K; 42 M; 100F; 45 P; 176S; 97 T; 7 W; 135I; 116V;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
 Residue Identity = 14% Matches = 1 Mismatches = 6
 Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXXKRDAIHQAFHNLNMPMDKIHETHIDY
1120 X 1130

6. US-08-121-713B-7 (1-7)
R25980 RPTP-beta amino acid sequence.

ID R52580 standard; protein; 2308 AA.
AC R52590;
DT 02-FEB-1995 (first entry)
DE RPTP-beta amino acid sequence.
KW Receptor protein tyrosine phosphatase; RPTP-beta; extracellular; CAM;
KW carbonic anhydrase; structural domain; cytoplasmic; phosphatase domains;
KW ligand; receptor-ligand complex; cell adhesion molecule; N-CAM; Ng-CAM;
KW growth factor; extracellular matrix component; enzymatic activity;
KW differentiation; metabolism; cell cycle control; neuronal function;
KW contractility; contact inhibition; viral receptor interactions;
KW inflammations; cellular transformation; cancer; type 2; NIDDM;
KW non-insulin dependant diabetes melitus.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Domain 33..301
FT /note= "Carbonic anhydrase domain"
FT Modified site 108
FT /note= "N-linked glycosylation site"
FT Modified site 134
FT /note= "N-linked glycosylation site"
FT Modified site 223
FT /note= "N-linked glycosylation site"
FT Modified site 232
FT /note= "N-linked glycosylation site"
FT Modified site 325
FT /note= "N-linked glycosylation site"
FT Modified site 381
FT /note= "N-linked glycosylation site"
FT Modified site 480
FT /note= "N-linked glycosylation site"
FT Modified site 497
FT /note= "N-linked glycosylation site"
FT Modified site 501
FT /note= "N-linked glycosylation site"
FT Modified site 552
FT /note= "N-linked glycosylation site"
FT Modified site 602
FT /note= "N-linked glycosylation site"
FT Modified site 629
FT /note= "N-linked glycosylation site"
FT Modified site 678
FT /note= "N-linked glycosylation site"
FT Modified site 782
FT /note= "N-linked glycosylation site"
FT Modified site 1017
FT /note= "N-linked glycosylation site"
FT Modified site 1050

/note= "N-linked glycosylation site"
FT Modified site 1082
FT /note= "N-linked glycosylation site"
FT Modified site 1122
FT /note= "N-linked glycosylation site"
FT Modified site 1457
FT /note= "N-linked glycosylation site"
FT Modified site 1562
FT /note= "N-linked glycosylation site"
FT Modified site 1618
FT /note= "N-linked glycosylation site"
FT Domain 1635..1662
FT /note= "Transmembrane domain"
FT Domain 1743..1984
FT /note= "DI phosphatase domain"
FT Domain 2041..2274
FT /note= "DII phosphatase domain"
FN WO9409037-A.
PD 28-APR-1994.
PF 14-OCT-1993; U09838.
PR 15-OCT-1992; US-961235.
PA (UWNY-) UNIV NEW YORK MEDICAL CENT.
PI Barnea G, Grumet MH, Margolis RU, Schlessinger J;
DR WPI; 94-151249/18.
PT New class of receptor protein tyrosine phosphatase molecule -
PT useful for identifying cpds which modulate the enzymatic activity
PT of the receptors
PS Claim 3; Fig 1; 6pp; English.
CC This sequence represents the receptor protein tyrosine phosphatase,
CC RPTP-beta. This protein comprises an extracellular carbonic
CC anhydrase structural domain and two cytoplasmic phosphatase domains.
CC RPTP-beta non-covalently binds a ligand to form a receptor-ligand
CC complex. The ligand is pref. a cell adhesion molecule (CAM), esp. N-
CC CAM or Ng-CAM, or a growth factor or extracellular matrix component.
CC The endogenous enzymatic activity of RPTP-beta may be increased or
CC decreased. This modulation regulates cellular functions comprising
CC differentiation, metabolism, cell cycle control or neuronal function.
CC The modulation pref. regulates cellular behaviour comprising
CC contractility or contact inhibition. It esp. regulates abnormal or
CC deleterious processes comprising viral receptor interactions, or
CC inflammations, cellular transformation to a cancerous state, or
CC development of type 2 NIDDM.
SQ Sequence 2308 AA;
SQ 144A; 47 R; 86 N; 135D; 0 B; 32 C; 96 Q; 166E; 0 Z; 119G; 67 H;
SQ 123I; 186L; 109K; 58 M; 88 F; 122P; 286S; 179T; 20 W; 74 Y; 159V;
SQ 12 Others;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXXPVMSQPQSVTDLEMPHYSTFAYFDTEV
700 710 X 720

7. US-08-121-713B-7 (1-7)
R67819 Acetyl CoA carboxylase.

ID R67819 standard; Protein; 2240 AA.
AC R67819;
DT 01-MAR-1995 (first entry)
DE Acetyl CoA carboxylase.
KW Acetyl CoA; carboxylase; lipid; fatty acid; biosynthesis;
OS metabolism; Brassica napus.
PN Brassica napus.
PD DE4317260-A.
PR 28-JUL-1994.
PF 24-MAY-1993; 317260.
PR 22-JAN-1993; DE-301694.
PR 24-MAY-1993; DE-317260.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Schell J, Schulte W, Toepfer R;
DR WPI; 94-236136/29.
DR N-PSDB; Q67676.
PT Acetyl CoA-carboxylase from plants, e.g. brassica napus - is
PT useful for modifying the oil and fatty acid prodn. in plants or
PT for conferring resistance to certain herbicides
PS Disclosure; Figure 7; 31pp; German.
CC Acetyl CoA carboxylase (ACC) is an enzyme which appears to be part of
CC the lipid storage system. ACC is end product regulated and is also
CC the limiting enzyme in de novo fatty acid biosynthesis. A changed
CC enzyme can interfere with both the lipid and fatty acid metabolism in
CC plants.
SQ Sequence 2240 AA;
SQ 177A; 134R; 81 N; 109D; 0 B; 34 C; 84 Q; 157E; 0 Z; 162G; 58 H;
SQ 121L; 220L; 114K; 62 M; 84 F; 108P; 162S; 98 T; 30 W; 65 Y; 180V;
Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
X X
XXXXPXX
DKIGSSLIQAQADVPTLPWSGSHVKIP
260 X 270 280

8. US-08-121-713B-7 (1-7)
R53257 Human collagen (Type V).
ID R53257 standard; Protein; 1838 AA.
AC R53257;
DT 12-JAN-1995 (first entry)
DE Human collagen (Type V).
KW Human collagen; alpha 1; V type collagen; placental mRNA.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..38
FT /label= signal peptide
FT Region 444..538
FT /note= "contains (Gly-X-Y) repeats"
FT Binding site 645..647
FT /label= RGD
FT /note= "cell adhesion motif"
FT Binding site 663..665
FT /label= RGD

/note= "cell adhesion motif"
FT Domain 897..933
FT /label= heparin binding domain
FT Region 1573..1838
FT /label= C-terminal region
FT /note= "contains 8 Cys residues"
PN J06105687-A.
PD 19-APR-1994.
PR 27-DEC-1991; 358300.
PR 27-DEC-1991; JP-358300.
PA (TAKI) TAKARA SHUZO CO LTD.
DR WPI; 94-163129/20.
DR N-PSDB; Q64556.
PT Human collagen V-type gene - is used for diagnosis of human
PT collagen V-type related diseases
PS Claim 1; Page 6-14; 19pp; Japanese.
CC This amino acid sequence of type V collagen contains several distinct
CC domains including a region comprising repeated (Gly-X-Y) motifs and a
CC central domain containing two copies of the Arg-Gly-Asp cell adhesion
CC motif. The cDNA sequence encoding type V collagen was isolated from a
CC human placental library and will be useful for diagnosis of diseases
CC related to type V collagen.
SQ Sequence 1838 AA;
SQ 92 A; 72 R; 34 N; 105D; 0 B; 12 C; 74 Q; 118E; 0 Z; 428G; 17 H;
SQ 52 I; 97 L; 98 K; 23 M; 39 F; 334P; 72 S; 70 T; 7 W; 40 Y; 54 V;
Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
X X
XXXXPXX
GRRGAGSDGARGMPGOTGPKGDRGFD
610 620 X 630

9. US-08-121-713B-7 (1-7)
R55273 Beta subunit of integrin cell surface receptor.
ID R55273 standard; Protein; 1822 AA.
AC R55273;
DT 31-JAN-1995 (first entry)
DE Beta subunit of integrin cell surface receptor.
KW Integrin; alpha; beta; subunit; glycoprotein; heterodimer;
KW transmembrane; extracellular matrix; cell signalling; cytoskeleton;
OS behaviour; signal transduction; receptor.
FH Key Location/Qualifiers
FT Peptide 28..42
FT /note= "N-terminal peptide."
FT Modified site 491
FT /note= "Potential N-linked glycosylation site."
FT Modified site 617
FT /note= "Potential N-linked glycosylation site."
FT Modified site 695
FT /note= "Potential N-linked glycosylation site."
FT Domain 711..733
FT /note= "Putative transmembrane domain."
FT Modified site 980

FT /note= "Potential N-linked glycosylation site."
FT Modified site 1593
FT /note= "Potential N-linked glycosylation site."
PN US5320942-A.
PD 14-JUN-1994.
PF 19-FEB-1987; 016552.
PR 19-FEB-1987; US-016552.
PR 04-JAN-1989; US-293384.
PR 01-OCT-1990; US-591105.
PA (KAJI/) KAJIJI S.
PA (QUAR/) QUARANTA V.
PI Kaji S, Quaranta V;
DR WP1; 94-191533/23.
DR N-PSDB; Q65674.
DR Diagnosing presence of abnormal epithelial tissue in vitro -
PT utilises monoclonal antibodies to alpha6 beta4 cell surface protein
PS Example 5; Figure 9; 34pp; English.
CC Integrins are heterodimers comprised of alpha and beta subunits, that
CC are non-covalently associated transmembrane glycoproteins. 11 alpha
CC chains and 6 beta chains have been recognised in man. Each alpha
CC subunit tends to associate with only one type of beta subunit but
CC there are several exceptions. Integrins mediate (in part) the
CC interaction of cells with the extracellular matrix, forming a link
CC between the extracellular matrix and the cytoskeleton. They may
CC transmit signals from the extracellular to the intracellular
CC environment, affecting cell behaviour. This sequence is the beta4
CC subunit of an alpha6 beta4 integrin.
SQ Sequence 1822 AA;
SQ 102A; 128R; 62 N; 102D; 0 B; 65 C; 88 Q; 123E; 0 Z; 135G; 46 H;
SQ 64 I; 175L; 63 K; 32 M; 58 F; 123P; 149S; 118T; 20 W; 50 Y; 119V;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X X
XXXXPXX
MAGFRPSPWARLLAALI
X 10

10. US-08-121-713B-7 (1-7)

R59751 Type II collagen.

ID R59751 standard; protein; 1418 AA.
AC R59751;
DT 14-FEB-1995 (first entry)
DE Type II collagen.
KW Collagen; triple helix; articular cartilage; collagenase;
KW degradation; monoclonal antibody; epitope; matrix;
KW metalloproteinase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..24
FT /label= Signal peptide.
PN W09414070-A.
PD 23-JUN-1994.
PF 06-DEC-1993; CA0522.
PR 04-DEC-1992; US-984123.

PA (SHRI-) SHRINERS HOSPITALS FOR CRIPPLED CHILDREN.
PI Hollander AP, Poole AR;
DR WP1; 94-234222/28.
PT Determn. of cartilage degradation - using a monoclonal antibody
PT to measure the amt. of unwound collagen or fragments in samples
PS Disclosure; Figure 1; 119pp; English.
CC Type II collagen constitutes the bulk of the fibrillar backbone of
CC cartilage matrix. It is composed of a tightly wound triple helix
CC which can only be cleaved by the metalloproteinase collagenase to
CC produce 3/4 and 1/4 length alpha chain fragments. The destruction of
CC articular cartilage is due, in part, to the degradation of collagen.
CC incapable of maintaining its helical structure at physiological
CC temperatures, collagenase-cleaved collagens unwind and become
CC susceptible to further degradation by other proteinases. By
CC producing monoclonal antibodies directed against epitopes which are
CC only revealed when collagen is unwound, the antibodies provide a
CC means of determining the degradation of cartilage in a biological
CC sample. The antibodies do not bind to native helical collagen.
CC Epitopes used in the production of such antibodies are described in
CC R59749, R59750 and R67742.

SQ Sequence 1418 AA;
SQ 129A; 71 R; 30 N; 54 D; 10 C; 58 Q; 74 E; 0 Z; 400G; 8 H;
SQ 30 I; 54 L; 64 K; 17 M; 23 F; 262P; 46 S; 40 T; 6 W; 9 Y; 33 V;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X X
XXXXPXX

GPPGAGQPPGLQGMPEGRGAAGIAGPK
670 680 X 690

11. US-08-121-713B-7 (1-7)
R60292 Varicella zoster virus IEPI75/IE62.

ID R60292 standard; Protein; 1310 AA.
AC R60292;
DT 21-FEB-1995 (first entry)
DE Varicella zoster virus IEPI75/IE62.
KW VZV; immediate early protein; IEPI75; IE62; phosphoprotein; human;
KW T cells; vaccine; secreted; cytotoxic T lymphocyte; virions.
OS Varicella-zoster virus.
PN W09414962-A.
PD 07-JUL-1994.
PF 17-DEC-1993; E03626.
PR 23-DEC-1992; GB-026768.
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PI Bollen A, Haumont M, Jacobs P, Massae M;
DR WP1; 94-234697/28.
DR N-PSDB; Q70362.
PT Pure varicella zoster virus immediate early protein 175 - and its
PT fusion proteins, useful in vaccines, also related vectors
PS Claim 2; Fig 1; 29pp; English.
CC This sequence represents the Varicella-zoster virus (VZV) immediate
CC early protein 175 (IEPI75). IEPI75 is encoded by an open reading
CC frame designated ORF62 and the protein itself is sometimes referred
CC to as IE62. The protein is a phosphoprotein with a relative

CC molecular weight of 175 kD. It is recognised by human T cells and
CC is thought to be an important immune target. This protein and
CC derivatives of it may be used in vaccines to treat or prevent VZV
CC diseases. Preferred derivatives include a secreted derivative in
CC which amino acids 226-257 and amino acids 648-733 have been deleted.
CC IEPI75 should induce cytotoxic T lymphocyte at an early stage,
CC preventing assembly of new virions and spreading to new cells.
SQ Sequence 1310 AA;
SQ 143A; 127R; 21 N; 87 D; 0 B; 12 C; 40 Q; 72 E; 0 Z; 120G; 25 H;
SQ 30 I; 97 L; 35 K; 14 M; 30 F; 148P; 120S; 61 T; 18 W; 20 Y; 90 V;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

TSGPVDPPAVITMPLDGPAPNGGFR
730 740 X 750

12. US-08-121-713B-7 (1-7)
R54841 HER4.

ID R54841 standard; Protein; 1308 AA.
AC R54841;
DT 11-JAN-1995 (first entry)
DE HER4.
KW Human epidermal growth factor receptor; HER; EGF; tyrosine kinase;
KW cancer; neuronal tissue; muscle tissue; neoplasm;
KW carcinoma; primer; probe; PCR.
OS Homo sapiens.

FH Key Location/Qualifiers

FT Region 1..1308
FT /note= "claim 12"
FT Region 26..1308
FT /note= "claim 12"
FT Region 1..1045
FT /note= "claim 12"
FT Region 26..1045
FT /note= "claim 12"
FT Region 772..1308
FT /note= "claim 12"
FT /note= "claim 12"
PN EP-599274-A.
PD 01-JUN-1994.
PF 23-NOV-1993; 118837.
PR 24-NOV-1992; US-981165.
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Culouscou J, Plowman GD, Shoyab M;
DR WPI; 94-169599/21.
DR N-PSDB; Q64896.
PT New recombinant nucleic acid expressing HER4 - a new receptor
PT tyrosine kinase expressed in some cancer cells, and related
PT vectors, antibodies, ligands etc, for diagnosis and treatment of
PT cancers.

PS Claim 12; Fig 1; 104pp; English.
CC HER4 is the fourth member of the EGFR-family of tyrosine kinases and
CC is expressed in some human cancers and in some tissues of neuronal
CC or muscle origin. HER4 polynucleotides, opt. labelled, are useful

CC in assays (e.g. of HER4 mRNA to detect certain neoplasms, esp.
CC breast carcinoma) and as primers in PCR or as probes.

SQ Sequence 1308 AA;
SQ 69 A; 68 R; 76 N; 58 D; 0 B; 58 C; 47 Q; 84 E; 0 Z; 91 G; 32 H;
SQ 68 I; 110I; 66 K; 29 M; 48 F; 90 P; 80 S; 76 T; 16 W; 53 Y; 78 V;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX
QCGNGFTSHDCIYYPWTGHSTLPQHAR
630 X 640

13. US-08-121-713B-7 (1-7)
R52701 Plasmid pASK60-Strep reading frame "b" translation

ID R52701 standard; Protein; 1277 AA.

AC R52701;
DT 11-JAN-1995 (first entry)
DE Plasmid pASK60-Strep reading frame "b" translation.
KW Streptavidin binding peptide; fusion protein; pASK60-Strep;
KW affinity chromatography; purification; peptide tag; detection.
OS Synthetic.

FH Key Location/Qualifiers

FT Region 1..1277
FT /note= "translated from reading frame 'b' of
FT pASK60-Strep; all X's correspond to nonsense
FT codons"
FT Peptide 443..463
FT /label= OmpA_signal_peptide
FN GB272698-A.
PD 25-MAY-1994.
PF 01-NOV-1993; 022501.
PR 03-NOV-1992; DE-237113.
PA (KUEH-) KUEHN KONSTRUKTION GMBH & CO KG KLAUS.
PA (BLOA-) INST BIOANALYTIK GEMEINNUETZIGE GMBH.
PI Schmidt T, Skerra A;
DR WPI; 94-153484/19.
DR N-PSDB; Q62676.

PT New fusion peptide(s) - have easily controlled binding properties
PT and are capable of binding to streptavidin
PS Example 5; Fig 8; 53pp; English.

CC Plasmid pASK60-Strep was produced starting from pASK40
CC (Biotechnology 9, 273-278, 1991) using site-directed mutagenesis
CC and PCR. The plasmid contains an improved set of unique restriction
CC sites, including two sites located directly at the 3' end of the
CC region coding for the OmpA signal peptide. The polylinker is
CC followed by a DNA sequence coding for a streptavidin binding
CC peptide. The plasmid pASK60-Strep is useful for the expression of
CC polypeptides fused to a streptavidin binding peptide. The
CC production of such fusion proteins can be specifically detected
CC using a streptavidin-alkaline phosphatase conjugate.

SQ Sequence 1277 AA;
SQ 122A; 84 R; 51 D; 0 B; 21 C; 56 Q; 54 E; 0 Z; 83 G; 26 H;
SQ 60 I; 144L; 49 K; 26 M; 55 F; 71 P; 100S; 79 T; 13 W; 28 Y; 82 V;
SQ 33 Others;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

INEAIPNDERDITMPVAMATTLKLLT
890 X 900 910

14. US-08-121-713B-7 (1-7)
R52702 Plasmid pASK60-Strep reading frame "c" translation

ID R52702 standard; Protein; 1277 AA.
AC R52702;
DT 11-JAN-1995 (first entry)
DE Plasmid pASK60-Strep reading frame "c" translation.
KW Streptavidin binding peptide; fusion protein; pASK60-Strep;
KW affinity chromatography; purification; peptide tag; detection.
OS Synthetic.
FH Key Location/Qualifiers
FT Region 1..1277
FT /note= "translated from reading frame 'c' of
FT pASK60-Strep; all X's correspond to nonsense
FT codons"
FT Peptide 482...491
FT /label= streptavidin-binding peptide
FN GB2272698-A.
PD 25-MAY-1994.
PF 01-NOV-1993; 022501.
PR 03-NOV-1992; DE-237113.
PA (KUEH-) KUEHN KONSTRUKTION GMBH & CO KG KLAUS.
PI (BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.
PI Schmidt T, Skerra A;
DR WPI; 94-153484/19.
DR N-PSDB; Q62676.
PT New fusion peptide(s) - have easily controlled binding properties
PT and are capable of binding to streptavidin
PS Example 5; Fig 8; 53pp; English.
CC Plasmid pASK60-Strep was produced starting from pASK40
CC (Biotechnology 9, 273-278, 1991) using site-directed mutagenesis
CC and PCR. The plasmid contains an improved set of unique restriction
CC sites, including two sites located directly at the 3' end of the
CC region coding for the OmpA signal peptide. The polylinker is
CC followed by a DNA sequence coding for a streptavidin binding
CC peptide. The plasmid pASK60-Strep is useful for the expression of
CC polypeptides fused to a streptavidin binding peptide. The
CC production of such fusion proteins can be specifically detected
CC using a streptavidin-alkaline phosphatase conjugate.
SQ Sequence 1277 AA;
SQ 87 A; 137R; 44 N; 23 D; 0 B; 38 C; 37 Q; 25 E; 0 Z; 54 G; 33 H;
SQ 60 I; 113L; 58 K; 22 M; 64 F; 95 P; 111S; 63 T; 51 W; 29 Y; 63 V;
SQ 70 Others;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX
|
XXMXNOXRYTMSQSMFVSLIRFFPAWXT
30 X 40 X 50

15. US-08-121-713B-7 (1-7)
R54074 CryET5.

ID R54074 standard; Protein; 1229 AA.
AC R54074;
DT 02-FEB-1995 (first entry)
DE CryET5.
KW cryET4; cryET5; Lepidoptera; lepidopteran insect; insecticidal;
KW toxin; insecticidal crystal protein; ICP.
OS Bacillus thuringiensis.
FN US322687-A.
PD 21-JUN-1994.
PF 29-JUL-1993; 100709.
PR 29-JUL-1993; US-100709.
PA (ECOG-) ECOGEN INC.
PI Donovan WP, Gonzalez JM, Jany CS, Tan Y;
DR WPI; 94-199503/24.
DR N-PSDB; Q64112.
PT Isolated cryET4 gene and Bacillus thuringiensis cultures
PT transformed with this gene - used in compans. against
PT lepidopteran insects.
PS Disclosure; Col 29-38; 51pp; English.
CC B. thuringiensis strain EG5847 exhibits insecticidal activity
CC against lepidopteran insects. Two novel toxin genes from B.
CC thuringiensis EG5847 designated cryET4 and cryET5 produce
CC insecticidal proteins with activity against a broad spectrum of
CC lepidopteran insects. The gene sequences are given in Q64111-12.
SQ Sequence 1229 AA;
SQ 74 A; 81 R; 89 N; 62 D; 0 B; 18 C; 47 Q; 100E; 0 Z; 77 G; 28 H;
SQ 70 I; 104L; 31 K; 9 M; 52 F; 55 P; 79 S; 97 T; 20 W; 59 Y; 77 V;

Initial Score = 7 Optimized Score = 7 Significance = 2.92
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

YSFLVGLWPSGRDPWEIFLHVQLI
90 X 100

> O <
O I O IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq7pir.res made by on Fri 19 May 95 8:52:16-PDT.

Query sequence being compared: US-08-121-713B-7 (1-7)
Number of sequences searched: 75511
Number of scores above cutoff: 4262

Results of the initial comparison of US-08-121-713B-7 (1-7) with:
Data bank : PIR 43, all entries

A log-linear plot showing the distribution of scores for various words. The y-axis is labeled 'SCORE 01' and 'STDEV 0' with a logarithmic scale from 0 to 100,000. The x-axis is labeled 'SCORE 01' and 'STDEV 0' with a linear scale from 0 to 7. Words are plotted as points, with some labeled with their frequency (e.g., N 100000, U 50000, M 10000, B 5000, E 1000, R 500, O 100, F 100, S 100, Q 100, D 100, N 100, C 100, S 100). A dashed line is drawn at y=10000.

PARAMETERS

Unitarity	1	K-tuple	2
Similarity matrix	1	Joining penalty	20
Mismatch penalty	1.00	Window size	6
Gap penalty	0.05		
Gap size penalty	0		
Cutoff score	0		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:

Mean	Median	Standard Deviation
2	1	3.26

Times:	CPU	Total Elapsed
	00:01:11.00	00:01:11.00

```
Number of residues:      2246834
Number of sequences searched: 75511
Number of scores above cutoff: 4262
```

[illegible]

The scores below are sorted by initial score. Significance is calculated based on initial score.

4262 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig.	Frame
1. A46105	polyprotein(NS1, NS3, NS5, NS	3415	7	7	1.53	0
2. B49132	faf=rat facets gene (alternat	2747	7	7	1.53	0
3. A49132	faf=rat facets gene (alternat	2711	7	7	1.53	0
4. S17796	inositol-trisphosphate recept	2701	7	7	1.53	0
5. S18198	notch protein homolog - rat	2531	7	7	1.53	0
6. S30446	fatty-acid synthase (EC 2.3.1	2505	7	7	1.53	0
7. A46299	disabled product (alternative	2411	7	7	1.53	0
8. S14568	microtubule-associated protei	1828	7	7	1.53	0

9.	S13507	microtubule-associated protein	1825	7	7	1.53
10.	A48298	sodium channel homolog - jell	1739	7	7	1.53
11.	A37491	o1 putative helicase/polyme	1737	7	7	1.53
12.	A48613	polyprotein-gag(CA, IN, NA, N	1603	7	7	1.53
13.	A46221	160K abdominal segment format	1533	7	7	1.53
14.	S43323	hypothetical protein - wheat	1470	7	7	1.53
15.	S47423	virus envelope protein - tran	1449	7	7	1.53
16.	S36126	neural cell adhesion molecule	1259	7	7	1.53
17.	S28764	neurocan - rat	1257	7	7	1.53
18.	S17655	neural cell adhesion molecule	1255	7	7	1.53
19.	S16948	hypothetical protein IRS-1 -	1235	7	7	1.53
20.	A46194	high-molecular-weight neurofi	1200	7	7	1.53
21.	S15362	glutamate receptor, metabotro	1199	7	7	1.53
22.	A39963	envelope glycoprotein G(envel	1148	7	7	1.53
23.	S39253	nitric-oxide synthase (EC 1.1	1147	7	7	1.53
24.	A40126	nitric-oxide synthase - rat	1147	7	7	1.53
25.	A36968	pl-like adhesion precursor - M	1144	7	7	1.53
26.	S23600	integrin alpha chain - rat (f	1106	7	7	1.53
27.	S38783	integrin alpha chain - rat (f	1106	7	7	1.53
28.	A47521	putative RNA-dependent RNA p	1057	7	7	1.53
29.	A46284	Ca(2+)-ATPase, LCA1 - Lycoper	1048	7	7	1.53
30.	JC2314	chitin synthase (EC 2.4.1.16)	1013	7	7	1.53
31.	A48821	intercellular signaling prote	1004	7	7	1.53
32.	JQ1165	Env protein - Visna virus	992	7	7	1.53
33.	S16385	CSF-1 receptor - rat	978	7	7	1.53
34.	JH0312	glutamate receptor GLUR5-1 ch	920	7	7	1.53
35.	JC2315	chitin synthase (EC 2.4.1.16)	916	7	7	1.53
36.	A44884	sex-specific germline RNA-bin	915	7	7	1.53
37.	A48613	pol polyprotein - myeloblasto	896	7	7	1.53
38.	E49613	pol polyprotein - myeloblasto	896	7	7	1.53
39.	A47521	putative capsid protein - g	890	7	7	1.53
40.	JH0311	glutamate receptor K3 chain p	888	7	7	1.53
41.	S13677	glutamate receptor chain K2 p	883	7	7	1.53
42.	JH0310	glutamate receptor K2 chain p	883	7	7	1.53
43.	PC2219	polypeptide - Hepatitis C vir	876	7	7	1.53
44.	S22338	glycogen phosphorylase (EC 2	850	7	7	1.53
45.	S47410	DNA polymerase - human hepat	845	7	7	1.53

```

1. US-08-121-713B-7 (1-7)
   A46105      polyprotein(NS1, NS3, NS5, NS2A, NS2B, NS4A, NS4B,
               A46105      #type complete
               polyprotein(NS1, NS3, NS2A, NS2B, NS4A, NS4B, C=small
               capsid protein, E=large envelope protein,
               (PR1M=membrane-anchored protein precursor) - Powassan virus
               #format name Powassan virus
               07-Apr-1994; #sequence_revision 07-Apr-1994; #text_change
               07-Apr-1994
               A46105
               A46105
               Mandl, C.W.; Holzmann, H.; Kunz, C.; Heinz, F.X.
               Virology (1993) 194:173-184
               Complete genomic sequence of Powassan virus: evaluation of
               genetic elements in tick-borne versus mosquito-borne
               flaviviruses.
               #cross-references MUID:93242744
               #accession A46105
               ##status preliminary
               ##molecule type genomic RNA

```

```

#residues      1-3415 ##label MAN
#Cross-references NCBIN:130654; NCBI:P130655
# #note
#length 3415 #molecular-weight 378568 #checksum 9967

SUMMARY
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.53
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

```

2. US-08-121-713B-7 (1-7)
 B49132 faf=rat facets gene (alternatively spliced, transcribed in brain)
 B49132 B49132 #type complete
 ENTRY faf=rat facets gene (alternatively spliced, transcript 1) -
 TITLE Drosophila
 ORGANISM #formal name Drosophila
 DATE 19-Dec-1993; #sequence revision 19-Dec-1993; #text_change
 16-Nov-1994

Fischer-Vize, J.A.; Rubin, G.M.; Lehmann, R.
 Development (1992) 116:985-1000
 The fat facets gene is required for Drosophila eye and embryo
 development.
 #cross-references PMID:93202020
 #contents isogenic st
 #accession B49132
 #status preliminary
 #molecule_type nucleic acid
 #residues 1-2747 #label FTS
 #cross-references NCBIN:127836; NCBIN:129008; NCBI:127839
 sequence inconsistent with the nucleotide translation
 sequence extracted from NCBI backbone
 #length 2747 #molecular-weight 307955 #checksum 2752
 SUMMARY
 SEQUENCE

Initial Score	=	7	Optimized Score	=	7	Significance	=	1.53
Residue Identity	=	14%	Matches	=	1	Mismatches	=	6
Gaps	=	0	Conservative Substitutions	=	0		=	0

X	X	
XXXXPXX		
AYFSHRSQFLPHCMPEDEMDWLITADRM		
440	450	460

3. US-08-121-713B-7 (1-7)
A49132 fat=fat facets gene (alternatively spliced, transc

ORGANISM Drosophila
DATE #formal name Drosophila
19-Dec-1993; #sequence_revision 19-Dec-1993; #text_change
19-Dec-1993
ACCESSIONS A49132
REFERENCE A49132
#authors Fischer-Vize, J.A.; Rubin, G.M.; Lehmann, R.
#journal Development (1992) 116:985-1000
#title The fat facets gene is required for Drosophila eye and embryo
development.
#cross-references MUID:93202020
#contents isogenic st
#accession A49132
#status preliminary
#molecule type nucleic acid
#residues 1-2711 ##label FIS
#cross-references NCBI:127836; NCBI:129008; NCBI:129029
#note sequence extracted from NCBI backbone
#length 2711 #molecular-weight 304063 #checksum 2362
SUMMARY
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.53
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

AYFSRHSQPLPHCPEDMDWLTADRM
440 450 X 460

4. US-08-121-713B-7 (1-7)
S17796 inositol-trisphosphate receptor type 2 - rat

ENTRY S17796 #type complete
TITLE inositol-trisphosphate receptor type 2 - rat
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993

ACCESSIONS S17796
REFERENCE S17796
#authors Sudhof, T.C.; Newton, C.L.; Archer III, B.T.; Ushkaryov,
Y.A.; Mignery, G.A.
#journal EMBO J. (1991) 10:3199-3206
#title Structure of a novel InsP(3) receptor.
#cross-references MUID:92007769
#accession S17796

##status preliminary
##residues 1-2701 ##label SUE
#cross-references EMBL:X61677

SUMMARY #length 2701 #molecular-weight 307056 #checksum 2707
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.53
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

EGDNIVGDKVILMPVNAVAGQLHASNV
180 X 190

5. US-08-121-713B-7 (1-7)
S18188 notch protein homolog - rat

ENTRY S18188 #type complete
TITLE notch protein homolog - rat
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 19-Feb-1994; #sequence_revision 19-Feb-1994; #text_change
19-Feb-1994

ACCESSIONS S18188
REFERENCE S18188
#authors Weinmaster, G.; Roberts, V.J.; Lemke, G.
#journal Development (1991) 113:199-205
#title A homolog of Drosophila Notch expressed during mammalian
development.
#cross-references MUID:92111383

#accession S18188
##status preliminary
##residues 1-2531 ##label WEI
#cross-references EMBL:X57405
SUMMARY #length 2531 #molecular-weight 270907 #checksum 2705
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.53
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

GOYCTEDVDECOLMPNACQNAGTCHNS
290 300 X 310

6. US-08-121-713B-7 (1-7)
S30446 fatty-acid synthase (EC 2.3.1.85) - rat

ENTRY S30446 #type complete
TITLE fatty-acid synthase (EC 2.3.1.85) - rat
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 08-Jun-1994; #sequence_revision 08-Jun-1994; #text_change
08-Jun-1994

ACCESSIONS S30446
REFERENCE S30446
#authors Beck, K.F.; Schreglmann, R.; Stathopoulos, I.; Klein, H.;
Hoch, J.; Schweizer, M.
#journal DNA Seq. (1992) 2:359-386
#title The fatty acid synthase (FAS) gene and its promoter in Rattus
norvegicus.

#accession S30446
##status preliminary
##residues 1-2505 ##label BEC
#cross-references EMBL:X62888

SUMMARY #length 2505 #molecular-weight 272647 #checksum 9219
SEQUENCE

```
Initial Score      = 7  Optimized Score = 7  Significance = 1.53
Residue Identity  = 14% Matches      = 1  Mismatches  = 6
Gaps              = 0  Conservative Substitutions = 0

      X      X
      XXXXPXX
      |
SMLNDIAATPTAAMPFRGYTVLGVGH
460 X      470

9. US-08-121-713B-7 (1-7)
A46299 disabled product (alternatively spliced) - Drosophila

ENTRY      A46299      #type complete
TITLE      disabled product (alternatively spliced) - Drosophila
ORGANISM   #formal name Drosophila
DATE       22-Sep-1993; #sequence_revision 22-Sep-1993; #text_change
17-Nov-1994
ACCESSIONS A46299
REFERENCE  Gertler, F.B.; Hill, K.K.; Clark, M.J.; Hoffmann, F.M.
            Genes Dev. (1993) 7:441-453
            Dosage-sensitive modifiers of Drosophila abl tyrosine kinase
            function: prospero, a regulator of axonal outgrowth, and
            disabled, a novel tyrosine kinase substrate.
            #cross-references NCBI:127163
            #accession MUID:93194063
            #status preliminary
            #molecule type nucleic acid
            #residues 1-2411 #label GER
            #cross-references NCBI:127163
            #note sequence extracted from NCBI backbone
            #note sequence not compared to nucleotide translation
SUMMARY    #length 2411 #molecular-weight 264046 #checksum 83
SEQUENCE

      X      X
      XXXXPXX
      |
TFVANFANFNDAPTPMTVSPVATVP
1110 X 1120 1130

8. US-08-121-713B-7 (1-7)
SI4568 microtubule-associated protein MAP2 - rat

ENTRY      SI4568      #type complete
TITLE      microtubule-associated protein MAP2 - rat
ORGANISM   #formal name Rattus norvegicus #common name Norway rat
DATE       21-Nov-1993; #sequence_revision 21-Nov-1993; #text_change
21-Nov-1993
ACCESSIONS SI4568
REFERENCE  Matus, A.; Doll, T.
            submitted to the EMBL Data Library, May 1990
```

```
#accession SI4568 preliminary
#status 1-1828 #label MAT
#residues 1-1828 #label MAT
#cross-references EMBL:X53455
SUMMARY    #length 1828 #molecular-weight 198982 #checksum 5694
SEQUENCE

Initial Score      = 7  Optimized Score = 7  Significance = 1.53
Residue Identity  = 14% Matches      = 1  Mismatches  = 6
Gaps              = 0  Conservative Substitutions = 0

      X      X
      XXXXPXX
      |
PLAPKDKQDWFTEMPVESKDEWGLAA
260 X 270 X 280

9. US-08-121-713B-7 (1-7)
SI3507 microtubule-associated protein MAP2 - rat

ENTRY      SI3507      #type complete
TITLE      microtubule-associated protein MAP2 - rat
ORGANISM   #formal name Rattus norvegicus #common name Norway rat
DATE       21-Nov-1993; #sequence_revision 21-Nov-1993; #text_change
21-Nov-1993
ACCESSIONS SI3507
REFERENCE  Marechal, D.; Delapierre, D.; Dresse, A.
            Arch. Int. Physiol. Biochim. (1988) 96:231-236
            Cloning and partial sequencing of a new rat brain specific
            cDNA
            #cross-references MUID:89334524
            #accession SI3507
            #status preliminary
            #residues 1-1825 #label MAR
            #cross-references EMBL:X54100
SUMMARY    #length 1825 #molecular-weight 198565 #checksum 5960
SEQUENCE

Initial Score      = 7  Optimized Score = 7  Significance = 1.53
Residue Identity  = 14% Matches      = 1  Mismatches  = 6
Gaps              = 0  Conservative Substitutions = 0

      X      X
      XXXXPXX
      |
PLAPKDKQDWFTEMPVESKDEWGLAA
260 X      270

10. US-08-121-713B-7 (1-7)
A48298 sodium channel homolog - jellyfish (Cyanea capilla)

ENTRY      A48298      #type complete
TITLE      sodium channel homolog - jellyfish (Cyanea capillata)
ORGANISM   #formal name Cyanea capillata
DATE       03-Jun-1994; #sequence_revision 03-Jun-1994; #text_change
03-Jun-1994
ACCESSIONS A48298
```


REFERENCE A48298
#authors Anderson, P.A.V.; Holman, M.A.; Greenberg, R.M.
#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:7419-7423
#title Deduced amino acid sequence of a putative sodium channel from the scyphozoan jellyfish Cyanea capillata.

#accession A48298
##status preliminary
##residues 1-1739 ##label AND
##cross-references GB:115445
SUMMARY #length 1739 #molecular-weight 197832 #checksum 3088
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.53
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

ASTAWNCPNYLCMPGAGSNPKNGYVS
330 X 340

11. US-08-121-713B-7 (1-7)
A37491 orf1 putative helicase/polymerase polyprotein - So
ENTRY orf1 putative helicase/polymerase polyprotein - Southampton
TITLE virus
ORGANISM #common name Southampton virus
DATE 03-Mar-1994; #sequence_revision 03-Mar-1994; #text_change
17-Nov-1994

ACCESSIONS A37491
REFERENCE A37491
#authors Lambden, P.R.; Caul, E.O.; Ashley, C.R.; Clarke, I.N.
#journal Science (1993) 259:516-519
#title Sequence and genome organization of a human small round-structured (Norwalk-like) virus.

#cross-references MUID:93142023
#contents small round-structured virus, SRSV, Norwalk virus,
Norwalk-like virus, serotype 3

#accession A37491
##status preliminary
##molecule_type genomic RNA
##residues 1-1737 ##label IAM
##cross-references NCBI:123456
##note sequence extracted from NCBI backbone
##note sequence not compared to nucleotide translation
SUMMARY #length 1737 #molecular-weight 193342 #checksum 6745
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.53
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

KEWFDAGLGCTMPPTTYERVKDSDPP
70 80 X 90

12. US-08-121-713B-7 (1-7)
A48613 polyprotein=gag(CA, IN, MA, NC, PR, GAG, POL, p10)

ENTRY A48613 #type complete
TITLE polyprotein=gag(CA, IN, MA, NC, PR, GAG, POL, p10) -
myeloblastosis-associated virus MAV
ORGANISM #formal name myeloblastosis-associated virus, MAV
DATE 21-Jan-1994; #sequence_revision 21-Jan-1994; #text_change
21-Jan-1994

ACCESSIONS A48613
REFERENCE A48613
#authors Joliot, V.; Boroughs, K.; Lasserre, F.; Crochet, J.;
Dambrine, G.; Smith, R.E.; Perbal, B.
#journal Virology (1993) 195:812-819
#title Pathogenic potential of myeloblastosis-associated virus:
implication of env proteins for osteopetrosis induction.

#cross-references MUID:9331743
#contents MAV2(O)/2
#accession A48613
##status preliminary
##molecule_type nucleic acid
##residues 1-1603 ##label JOL
##cross-references NCBI:135486; NCBI:135487
##note sequence extracted from NCBI backbone
SUMMARY #length 1603 #molecular-weight 173689 #checksum 1201
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 1.53
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

REELASTGPPVWVMPVVIKTEGPAWTP
230 X 240 X 250

13. US-08-121-713B-7 (1-7)
A46221 160K abdominal segment formation protein Pumilio -

ENTRY A46221 #type complete
TITLE 160K abdominal segment formation protein Pumilio - Drosophila
ORGANISM #formal name Drosophila
DATE 21-Sep-1993; #sequence_revision 21-Sep-1993; #text_change
17-Nov-1994

ACCESSIONS A46221
REFERENCE A46221
#authors Barker, D.D.; Wang, C.; Moore, J.; Dickinson, L.K.; Lehmann,
R.
#journal Genes Dev. (1992) 6:2312-2326
#title Pumilio is essential for function but not for distribution of
the Drosophila abdominal determinant Nanos.

#cross-references MUID:93093466
#contents embryo
#accession A46221
##status preliminary
##molecule_type nucleic acid

```
##residues      1-1533 ##label BAR
##cross-references NCBI:120203; NCBI:120204
##note          sequence extracted from NCBI backbone
SUMMARY        #length 1533 #molecular-weight 157453 #checksum 5038
SEQUENCE
```

```
Initial Score      = 7 Optimized Score = 7 Significance = 1.53
Residue Identity   = 14% Matches      = 1 Mismatches  = 6
Gaps               = 0 Conservative Substitutions = 0
```

```
X      X
XXXXPXX
```

```
HHGGAMHFGMGMPKQQLGPFPGAGG
310 X 320 X 330
```

```
14. US-08-121-713B-7 (1-7)
S45323 hypothetical protein - wheat spindle streak mosaic

ENTRY          S45323 #type complete
TITLE          hypothetical protein - wheat spindle streak mosaic virus
ORGANISM       #formal_name wheat spindle streak mosaic virus
DATE           20-Oct-1994; #sequence_revision 20-Oct-1994; #text_change
20-Oct-1994
ACCESSIONS     S45323
REFERENCE      S45323
#authors       Sohn, A.; Schenk, P.; Signoret, P.A.; Schmitz, G.; Schell,
J.; Steinbiss, H.H.
#journal       Arch. Virol. (1994) 135:279-292
#title         Sequence analysis of the 3'-terminal half of RNA 1 of wheat
spindle streak mosaic virus.
```

```
#accession      S45323
#status         preliminary
#residues       1-1470 ##label SOH
##cross-references EMBL:X73883
```

```
SUMMARY        #length 1470 #molecular-weight 163878 #checksum 2907
SEQUENCE
```

```
Initial Score      = 7 Optimized Score = 7 Significance = 1.53
Residue Identity   = 14% Matches      = 1 Mismatches  = 6
Gaps               = 0 Conservative Substitutions = 0
```

```
X      X
XXXXPXX
```

```
KSSLNSSLNTMPVSEGEILKEVM
180 190 X 200
```

```
15. US-08-121-713B-7 (1-7)
S47423 virus envelope protein - transmissible gastroenteritis virus

ENTRY          S47423 #type complete
TITLE          virus envelope protein - transmissible gastroenteritis virus
ORGANISM       #formal_name transmissible gastroenteritis virus
DATE           23-Nov-1994; #sequence_revision 23-Nov-1994; #text_change
23-Nov-1994
ACCESSIONS     S47423
REFERENCE      S47422
```

```
#authors        Chen, C.; Cavanagh, D.; Britton, P.
#submission     submitted to the EMBL Data Library, August 1994
#accession      S47423
#status         preliminary
#residues       1-1449 ##label CHE
```

```
##cross-references EMBL:Z35758
SUMMARY        #length 1449 #molecular-weight 160118 #checksum 3476
SEQUENCE
```

```
Initial Score      = 7 Optimized Score = 7 Significance = 1.53
Residue Identity   = 14% Matches      = 1 Mismatches  = 6
Gaps               = 0 Conservative Substitutions = 0
```

```
X      X
XXXXPXX
```

```
NSDVLGDFPTVQFWFNCIRNNSNDL
50 60 X 70
```

```
> O <
OI IO IntelliGenetics
> O <
```

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq7spt.res made by on Fri 19 May 95 8:55:49-PDT.

Query sequence being compared:US-08-121-713B-7 (1-7)
Number of sequences searched: 43470
Number of scores above cutoff: 4963

Results of the initial comparison of US-08-121-713B-7 (1-7) with:
Data bank : Swiss-Prot 31, all entries

```
100000-
-
N
U50000-
M
B *
E
R
O
F10000-
-
S
E 5000-
Q
U
N
E
C
S 1000-
-
500-
-
*
```

Cut-off raised to 7.
Cut-off raised to 7.
Cut-off raised to 7.
Cut-off raised to 7.
Cut-off raised to 7.
Cut-off raised to 7.

The scores below are sorted by initial score. Significance is calculated based on initial score.

4963 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig. Frame
1. YHP9 YEAST	HYPOTHETICAL 433.2 KD PROTEIN	3744	7	7	1.48
2. ZFH2 DROME	ZINC-FINGER PROTEIN 2 (ZINC-F	3005	7	7	1.48
3. ZEP1_HUMAN	ZINC FINGER PROTEIN 40 (HUMAN	2717	7	7	1.48
4. YLJ6_CAEEL	HYPOTHETICAL 272.0 KD PROTEIN	2329	7	7	1.48
5. YLJ1_CAEEL	HYPOTHETICAL 223.8 KD PROTEIN	2027	7	7	1.48
6. YHD0 YEAST	HYPOTHETICAL 210.4 KD PROTEIN	1868	7	7	1.48
7. ZEP2_HUMAN	HUMAN IMMUNODEFICIENCY VIRUS	1833	7	7	1.48
8. YM68_CAEEL	HYPOTHETICAL 208.3 KD PROTEIN	1822	7	7	1.48
9. YQ25_CAEEL	HYPOTHETICAL 202.6 KD PROTEIN	1799	7	7	1.48
10. YK84 YEAST	HYPOTHETICAL 203.3 KD PROTEIN	1764	7	7	1.48
11. YI97 YEAST	HYPOTHETICAL 197.5 KD PROTEIN	1758	7	7	1.48
12. YEW2 YEAST	HYPOTHETICAL 195.4 KD PROTEIN	1753	7	7	1.48
13. YK11_CAEEL	HYPOTHETICAL 195.2 KD PROTEIN	1683	7	7	1.48
14. YL54_CAEEL	HYPOTHETICAL 176.0 KD PROTEIN	1609	7	7	1.48
15. YIR3 YEAST	POTATIVE MEMBRANE GLYCOPROTEI	1549	7	7	1.48
16. YU64 YEAST	POTATIVE MEMBRANE GLYCOPROTEI	1541	7	7	1.48
17. YHY6 YEAST	HYPOTHETICAL 175.8 KD PROTEIN	1549	7	7	1.48
18. YHM4 YEAST	HYPOTHETICAL 171.7 KD PROTEIN	1522	7	7	1.48
19. YEZ4 YEAST	HYPOTHETICAL 168.2 KD PROTEIN	1468	7	7	1.48
20. YEC2 YEAST	HYPOTHETICAL 165.7 KD PROTEIN	1459	7	7	1.48
21. YN81_CAEEL	HYPOTHETICAL 155.8 KD PROTEIN	1439	7	7	1.48
22. YI33 YEAST	PROBABLE ATP-DEPENDENT PERMEA	1411	7	7	1.48
23. YMS5_CAEEL	HYPOTHETICAL 161.8 KD PROTEIN	1409	7	7	1.48
24. YTP9 YEAST	HYPOTHETICAL 156.9 KD PROTEIN	1375	7	7	1.48
25. YH00 YEAST	HYPOTHETICAL 149.7 KD PROTEIN	1345	7	7	1.48
26. YK09 YEAST	MYOSIN HEAVY CHAIN HOMOLOG YK	1271	7	7	1.48
27. YNCA_CAEEL	HYPOTHETICAL 142.5 KD PROTEIN	1254	7	7	1.48
28. YMF6_CAEEL	HYPOTHETICAL 139.9 KD PROTEIN	1247	7	7	1.48
29. YQ05_CAEEL	HYPOTHETICAL 134.9 KD PROTEIN	1232	7	7	1.48
30. YCX1 YEAST	HYPOTHETICAL 143.8 KD PROTEIN	1226	7	7	1.48
31. YMA1_CAEEL	PROBABLE INTEGRIN ALPHA CHAIN	1226	7	7	1.48
32. YS12 YEAST	HYPOTHETICAL 137.7 KD PROTEIN	1224	7	7	1.48
33. YMH5_CAEEL	HYPOTHETICAL 136.3 KD PROTEIN	1222	7	7	1.48
34. YK83 YEAST	PROBABLE ATP-DEPENDENT PERMEA	1218	7	7	1.48
35. YED1 YEAST	PROBABLE EI-E2 ATPASE YEL031W	1215	7	7	1.48
36. YNM5_CAEEL	HYPOTHETICAL 139.4 KD PROTEIN	1213	7	7	1.48
37. YSH1_ECOLI	HYPOTHETICAL 138.1 KD PROTEIN	1210	7	7	1.48
38. YI08_CAEEL	HYPOTHETICAL 138.6 KD PROTEIN	1207	7	7	1.48
39. YK76 YEAST	HYPOTHETICAL 137.5 KD PROTEIN	1195	7	7	1.48
40. ZN91_HUMAN	ZINC FINGER PROTEIN 91 (ZINC	1191	7	7	1.48
41. YK08 YEAST	POTATIVE 128.2 KD TRANSCRIPTI	1170	7	7	1.48
42. YK82 YEAST	HYPOTHETICAL 122.2 KD PROTEIN	1169	7	7	1.48
43. YK62 YEAST	HYPOTHETICAL 133.3 KD PROTEIN	1157	7	7	1.48
44. YI68 YEAST	PROBABLE EI-E2 ATPASE YI048W	1151	7	7	1.48

45. YHC3_YEAST HYPOTHETICAL 130.0 KD PROTEIN 1146 7 1.48 0

1. US-08-121-713B-7 (1-7)
YHP9_YEAST HYPOTHETICAL 433.2 KD PROTEIN IN HXT5-CDC12 INTERG

ID YHP9_YEAST STANDARD; PRT; 3744 AA.
AC P38811;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 433.2 KD PROTEIN IN HXT5-CDC12 INTERGENIC REGION.
GN YHR099W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RM 94378003
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,
RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
RA NHAN M., RIFKIN L., RILES L., ST.PETER H., TREVASKIS E., VAUGHAN K.,
RA VIGNATI D., WILCOX L., WOLDMAN P., WATERSTON R., WILSON R.,
RA VAUDIN M.;
RL SCIENCE 265:2077-2082(1994).
DR EMBL; U00060; SCH9332.
DR PIR; S46715; S46715.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 3744 AA; 433171 MW; 20943107 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXX

MRKVLPSDDHLMFPQPKEDINDSPD
520 530 X 540

2. US-08-121-713B-7 (1-7)
ZFH2_DROME ZINC-FINGER PROTEIN 2 (ZINC-FINGER HOMEODOMAIN PRO

ID ZFH2_DROME STANDARD; PRT; 3005 AA.
AC P28167;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE ZINC-FINGER PROTEIN 2 (ZINC-FINGER HOMEODOMAIN PROTEIN 2).
GN ZFH-2.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RM 92001539
RA FORTINI M.E., LAI Z., RUBIN G.M.;
RL MECH. DEV. 34:113-122(1991).

CC -!- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
CC NERVOUS SYSTEM.

CC -!- TISSUE SPECIFICITY: LARGELY RESTRICTED TO THE CNS OF LATE EMBRYO.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).

CC -!- SIMILARITY: CONTAINS THREE HOMEBOX DOMAINS.

DR EMBL; M63450; DMZF2.

DR PIR; S27817; S27817.

DR PIR; S33642; S33642.

DR HSSP; P02836; LHDD.

KW FLYBASE; FBGN0004607; ZFH2.

KW ZINC-FINGER; METAL-BINDING; DNA-BINDING; HOMEBOX; NUCLEAR PROTEIN;
KW REPEAT...

FT ZN_FING 133 156 C2H2-TYPE.

FT ZN_FING 559 582 C2H2-TYPE.

FT ZN_FING 614 638 C2H2-TYPE.

FT ZN_FING 732 756 C2H2-TYPE.

FT ZN_FING 897 916 C2H2-TYPE (DEGENERATE).

FT ZN_FING 940 964 C2H2-TYPE.

FT ZN_FING 999 1023 C2H2-TYPE.

FT ZN_FING 1074 1098 C2H2-TYPE.

FT ZN_FING 1210 1233 C2H2-TYPE.

FT ZN_FING 1341 1365 C2H2-TYPE.

FT ZN_FING 1438 1462 C2H2-TYPE.

FT ZN_FING 1477 1500 C2H2-TYPE (DEGENERATE).

FT ZN_FING 1513 1535 C2H2-TYPE.

FT ZN_FING 1541 1564 C2H2-TYPE.

FT DNA_BIND 1797 1856 HOMEBOX 1.

FT DNA_BIND 2154 2213 HOMEBOX 2.

FT ZN_FING 2234 2256 C2H2-TYPE.

FT ZN_FING 2371 2393 C2H2-TYPE.

FT DNA_BIND 2760 2819 HOMEBOX 3.

SQ SEQUENCE 3005 AA; 332056 MW; 22430417 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXX

GTGNTASHSSGSFMEVALADLAYNQ
790 X 800

3. US-08-121-713B-7 (1-7)
ZEP1_HUMAN ZINC FINGER PROTEIN 40 (HUMAN IMMUNODEFICIENCY VIR

ID ZEP1_HUMAN STANDARD; PRT; 2717 AA.
AC P15822;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ZINC FINGER PROTEIN 40 (HUMAN IMMUNODEFICIENCY VIRUS TYPE I ENHANCER-
DE BINDING PROTEIN 1) (HIV-EPI) (MAJOR HISTOCOMPATIBILITY COMPLEX BINDING
DE PROTEIN 1) (MBP-1) (POSITIVE REGULATORY DOMAIN II BINDING FACTOR 1)
DE (PRDII-BE1)
GN HIVEPI OR ZNF40.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.

[1] SEQUENCE FROM N.A.
 RP 90169514
 RA FAN C.M., MANIATIS T.;
 RL GENES DEV. 4:29-42(1990).
 [2] STRUCTURE BY NMR OF 2113-2142.
 RP 91064333
 RA OMICHINSKI J.G., CLORE G.M., APPELLA E., SAKAGUCHI K.,
 RA GRONENBORN A.M.;
 RL BIOCHEMISTRY 29:9324-9334(1990).
 [3] STRUCTURE BY NMR OF 2087-2142.
 RP 92232684
 RA OMICHINSKI J.G., CLORE G.M., ROBIEN M., SAKAGUCHI K., APPELLA E.,
 RA GRONENBORN A.M.;
 RL BIOCHEMISTRY 31:3907-3917(1992).
 CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
 CC 5'-GGGACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
 CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
 CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
 CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
 CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
 CC IN T-CELL ACTIVATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY SEPARATED
 CC AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH ZINC-FINGER
 CC IN-BETWEEN.
 CC -1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
 CC -1- SIMILARITY: 70% BETWEEN THE TWO ZINC-FINGER DOMAINS.
 CC -1- SIMILARITY: STRONG, TO HIV-EP2.
 DR EMBL; X51435; HS2FPBF1.
 DR PIR; A34203; A34203.
 DR PDB; 3ZNF; 15-JAN-92.
 DR PDB; 4ZNF; 15-JAN-92.
 DR PDB; 1BBO; 31-OCT-93.
 DR TRANSFAC; T00497; -.
 DR MIM; 194540; 11TH EDITION.
 DR PROSITE; PS00028; ZINC FINGER C2H2.
 KW TRANSCRIPTION REGULATION; ZINC-FINGER; METAL-BINDING; DNA-BINDING;
 KW NUCLEAR PROTEIN; 3D-STRUCTURE.
 FT DOMAIN 406 456
 FT ZN FING 406 428
 FT ZN FING 434 456
 FT ZN FING 953 981
 FT DOMAIN 2087 2139
 FT ZN FING 2087 2109
 FT ZN FING 2115 2139
 FT STRAND 2088 2088
 FT TURN 2090 2092
 FT STRAND 2093 2095
 FT HELIX 2099 2108
 FT TURN 2109 2109
 FT STRAND 2115 2116
 FT STRAND 2123 2124
 FT HELIX 2127 2135
 SQ SEQUENCE 2717 AA; 297217 MW; 19898490 CN;
 Initial Score = 7 Optimized Score = 7 Significance = 1.48
 Residue Identity = 25% Matches = 1 Mismatches = 3

Gaps = 0 Conservative Substitutions = 0
 XXXXPXX
 MPTKQIHPRNLRD
 X X 10
 4. US-08-121-713B-7 (1-7)
 YLJ6_CAEL HYPOTHETICAL 272.0 KD PROTEIN C50C3.6 IN CHROMOSOM
 ID YLJ6 CAEL STANDARD; PRT; 2329 AA.
 AC P34369;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 272.0 KD PROTEIN C50C3.6 IN CHROMOSOME III.
 GN C50C3.6.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RM 94150718
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
 RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
 RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
 RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WOHLDMAN P.;
 RL NATURE 368:32-38(1994).
 DR EMBL; L14433; CEC50C3.
 DR PIR; S44625; S44625.
 DR WORMPEP; C50C3.6; CE00122.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 2329 AA; 272025 MW; 19536537 CN;
 Initial Score = 7 Optimized Score = 7 Significance = 1.48
 Residue Identity = 14% Matches = 1 Mismatches = 6
 Gaps = 0 Conservative Substitutions = 0
 XXXXPXX
 KKKFGMSDTQKEEMPEHVRKVIKRDHG
 40 50 X 60
 5. US-08-121-713B-7 (1-7)
 YNJ1_CAEL HYPOTHETICAL 223.8 KD PROTEIN R10E11.1 IN CHROMOSO
 ID YNJ1 CAEL STANDARD; PRT; 2027 AA.
 AC P34545;
 DT 01-FEB-1994 (REL. 28, CREATED)

DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 223.8 KD PROTEIN R10E11.1 IN CHROMOSOME III.
GN R10E11.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RM 94150718
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKIN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SWALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
EL NATURE 368:32-38 (1994).
CC -!- SIMILARITY: CONTAINS A COPY OF THE BROMODOMAIN.
DR EMBL; Z29095; CER10E11.
DR PIR; S40713; S40713.
DR HSP; P19999; ICLG.
DR WORMPEP; R10E11.1; CE01049.
DR PROSITE; PS00633; BROMODOMAIN.
KW HYPOTHETICAL PROTEIN; BROMODOMAIN.
FT DOMAIN 886 948 BROMODOMAIN.
SQ SEQUENCE 2027 AA; 223840 MW; 18596053 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX
|
GQPMGRGAMNGAMPGRSGPMTQGRPG
230 240 X 250

6. US-08-121-713B-7 (1-7)
YHDO_YEAST HYPOTHETICAL 210.4 KD PROTEIN IN GUT1-RIM1 INTERGE
ID YHDO YEAST STANDARD; PRT; 1868 AA.
AC P38737;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 210.4 KD PROTEIN IN GUT1-RIM1 INTERGENIC REGION.
GN YHLO30W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RM 94378003
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,

DU Z., FAVELLO A., FULTON L., GATTONG S., GEISEL C., KIRSTEN J.,
RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
RA NHAN M., RIFKIN L., RILES L., ST.PETER H., TREVASKIS E., VAUGHAN K.,
RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
RA VAUDIN M.;
RL SCIENCE 265:2077-2082 (1994).
DR EMBL; U11583; SCH9196.
DR PIR; S48938; S48938.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 695 715 POTENTIAL.
FT TRANSMEM 747 767 POTENTIAL.
SQ SEQUENCE 1868 AA; 210430 MW; 17651645 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX
|
TNLSDDAAFMKRFMPYENEEFITFL
280 X 290

7. US-08-121-713B-7 (1-7)
ZEP2_HUMAN HUMAN IMMUNODEFICIENCY VIRUS TYPE I ENHANCER-BINDI
ID ZEP2 HUMAN STANDARD; PRT; 1833 AA.
AC P31629;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HUMAN IMMUNODEFICIENCY VIRUS TYPE I ENHANCER-BINDING PROTEIN 2
DE (HIV-EF2).
GN HIVEP2.
OS HOMO SAPIENS (HUMAN).
OC EDKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RM 91217105
RA NOMURA N., ZHAO M.-J., NAGASE T., MAEKAWA T., ISHIZAKI R., TABATA S.,
RA ISHII S.;
RL J. BIOL. CHEM. 266:8590-8594 (1991).
CC -!- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
CC 5'-GGGACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
CC IN T-CELL ACTIVATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
CC -!- SIMILARITY: TO THE HIV-EF1.
DR EMBL; M60119; HSEP2AA.
DR PIR; A39829; WMHUE2.
DR HSSP; P15822; 1BBO.
DR TRANSFAC; T00939; -.
DR MIM; 143054; 11TH EDITION.

DR PROSITE; PS00028; ZINC_FINGER_C2H2.
KW TRANSCRIPTION REGULATION; ZINC-FINGER; METAL-BINDING; DNA-BINDING;
FT DOMAIN 324 330 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 337 369 SER-RICH.
FT DOMAIN 1286 1310 ASP/GLU-RICH (ACIDIC).
FT ZN FING 1186 1208 C2H2-TYPE.
FT ZN FING 1214 1238 C2H2-TYPE.
SQ SEQUENCE 1833 AA; 202128 MW; 17625550 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

QYPTVMVHLPAQOPWQHHPFA
600 X 610 X 620

9. US-08-121-713B-7 (1-7)
YM68_CAEEL HYPOTHETICAL 208.3 KD PROTEIN K12H4.8 IN CHROMOSOM

ID YM68 CAEEL STANDARD; PRT; 1822 AA.

AC P34529;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 208.3 KD PROTEIN K12H4.8 IN CHROMOSOME III.
GN K12H4.8.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RM 94150718
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SOLSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RM 94150718
DR EMBL; L16621; CEZK688.
DR PIR; S44920; S44920.
DR WORMPEP; K12H4.8; CE00273.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1822 AA; 208291 MW; 15724531 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

QYPTVMVHLPAQOPWQHHPFA
600 X 610 X 620

10. US-08-121-713B-7 (1-7)
YK4_YEAST HYPOTHETICAL 203.3 KD PROTEIN IN PUT3-CCE1 INTERGE

ID YK4_YEAST STANDARD; PRT; 1764 AA.

AC P34241; P34242;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)

X X
XXXXPXX

DRGPGENDRNVMFDRVAGPRIINAI
100 110

Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX
DSIKTTAVFRLGFWAAWRTAQVWEK
280 290 X 300

9. US-08-121-713B-7 (1-7)
Y025_CAEEL HYPOTHETICAL 202.6 KD PROTEIN ZK688.5 IN CHROMOSOM

ID Y025 CAEEL STANDARD; PRT; 1799 AA.

AC P34675;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 202.6 KD PROTEIN ZK688.5 IN CHROMOSOME III.
GN ZK688.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RM 94150718
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SOLSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RM 94150718
DR EMBL; L16621; CEZK688.
DR PIR; S44920; S44920.
DR WORMPEP; ZK688.5; CE00463.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1799 AA; 202641 MW; 15440747 CN;

Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXPXX

DRGPGENDRNVMFDRVAGPRIINAI
100 110

10. US-08-121-713B-7 (1-7)
YK4_YEAST HYPOTHETICAL 203.3 KD PROTEIN IN PUT3-CCE1 INTERGE

ID YK4_YEAST STANDARD; PRT; 1764 AA.

AC P34241; P34242;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)

X X
XXXXPXX

DRGPGENDRNVMFDRVAGPRIINAI
100 110

DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 203.3 KD PROTEIN IN PUT3-CE1 INTERGENIC REGION.
GN YKL014C.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RW 94205264
RA WIEMANN S., VOSS H., SCHWAGER C., RUPP T., STEGEMANN J.,
RA ZIMMERMANN J., GROTHUES D., SENSEN C., ERFLE H., HEWITT N.,
RA BANREVI A., ANSORGE W.;
RA BANREVI A., ANSORGE W.;
RL YEAST 9:1343-1348(1993).
RN [2]
RP REVISIONS.
RA WIEMANN S., VOSS H., SCHWAGER C., RUPP T., STEGEMANN J.,
RA ZIMMERMANN J., GROTHUES D., SENSEN C., ERFLE H., HEWITT N.,
RA BANREVI A., ANSORGE W.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X74152; SCPOKGA.
DR EMBL; Z28014; SCYKLO14C.
DR PIR; S37827; S37827.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1764 AA; 203286 MW; 16178809 CN;
Initial Score = 7 Optimized Score = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
X X
XXXXPXX
|
IHNKLFNTIRLFKPEDTTLQLGLIK
340 X 350
11. US-08-121-713B-7 (1-7)
YIR7 YEAST HYPOTHETICAL 197.5 KD PROTEIN IN SUC2 5'REGION.
ID YIR7 YEAST STANDARD; PRT; 1758 AA.
AC P40434;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 197.5 KD PROTEIN IN SUC2 5'REGION.
GN YIL177C.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA WALSH S.V., WHITEHEAD S.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO HYPOTHETICAL PROTEIN IN SUBTELOMERIC Y'REPEAT
(Z46921).
DR EMBL; Z47047; SCCHRIX.

KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1758 AA; 197511 MW; 15509879 CN;
Initial Score = 7 Optimized Score = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
X X
XXXXPXX
|
ICALGNSYDAFNHDFMDVGVGFEDPNQ
450 X 460 X 470
12. US-08-121-713B-7 (1-7)
YEW2 YEAST HYPOTHETICAL 195.4 KD PROTEIN IN RPS26B-GLC7 INTER
ID YEW2 YEAST STANDARD; PRT; 1753 AA.
AC P32634;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 195.4 KD PROTEIN IN RPS26B-GLC7 INTERGENIC REGION.
GN YER132C OR SVGP-ORF50.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA MULLIGAN J.T., DIETRICH F.S., HENNESSEY K.M., SEHL P., KOMP C.,
RA WEI Y., TAYLOR P., NAKAHARA K., ROBERTS D., DAVIS R.W.;
RL SUBMITTED (FEB-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGHEN R., OEFNER P., OH C.,
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO S.POMBE RAL2.
DR EMBL; L11120; SCYGP4.
DR PIR; S30855; S30855.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1753 AA; 195382 MW; 16043187 CN;
Initial Score = 7 Optimized Score = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
X X
XXXXPXX
|
SKGSKILVNIDGMPFISLSHDSNFSVL
250 X 260 270

13. US-08-121-713B-7 (1-7)
YK11 YEAST HYPOTHETICAL 195.2 KD PROTEIN IN GCN3-DAL80 INTERG
ID YK11 YEAST STANDARD; PRT; 1683 AA.
AC P36126;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 195.2 KD PROTEIN IN GCN3-DAL80 INTERGENIC REGION.
GN YK031C.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA URRESTARAZU L.A., JAUNIAUX J.-C.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Z28256; SCYKRO31C.
DR PIR; S38103; S38103.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1683 AA; 195203 MW; 14046134 CN;
Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
X X
XXXXPXX
KQWSSIIKMTSTPWSKNRFGSFAP 670
660
14. US-08-121-713B-7 (1-7)
YL54 CAEEL HYPOTHETICAL 176.0 KD PROTEIN F44E2.4 IN CHROMOSOM
ID YL54 CAEEL STANDARD; PRT; 1609 AA.
AC F34434;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 176.0 KD PROTEIN F44E2.4 IN CHROMOSOME III.
F44E2.4.
GN CAENORHABDITIS ELEGANS.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOLELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RM 94150718
RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DUBBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SNALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RL NATURE 368:32-38 (1994).

DR EMBL; L23646; CEF44E2.
DR PIR; S44821; S44821.
DR WORMPEF; F44E2.4; CE00182.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1609 AA; 175966 MW; 13272801 CN;
Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
X X
XXXXPXX
TTYQECVGNQKGMPEDEGHSWGEVEV
120 130 X 140
15. US-08-121-713B-7 (1-7)
YIR3 YEAST PUTATIVE MEMBRANE GLYCOPROTEIN IN SUC2 5'REGION PR
ID YIR3 YEAST STANDARD; PRT; 1549 AA.
AC P40438;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE PUTATIVE MEMBRANE GLYCOPROTEIN IN SUC2 5'REGION PRECURSOR.
GN YI1173W.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA GENTILES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA WALSH S.V., WHITEHEAD S.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: ALMOST IDENTICAL TO YEAST HRC1549. ALSO STRONG, TO
CC YEAST PEPI.
DR EMBL; Z47047; SCCHRIX.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1549 MEMBRANE GLYCOPROTEIN YI1173W.
FT TRANSMEM 1370 1390 POTENTIAL.
FT CARBOHYD 479 479 POTENTIAL.
FT CARBOHYD 769 769 POTENTIAL.
FT CARBOHYD 986 986 POTENTIAL.
FT CARBOHYD 1418 1418 POTENTIAL.
SQ SEQUENCE 1549 AA; 174426 MW; 12283501 CN;
Initial Score = 7 Optimized Score = 7 Significance = 1.48
Residue Identity = 14% Matches = 1 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
X X
XXXXPXX
FMIAKLPEARQLGMPDLDFSAKAQDTFI

Listing for Mary Hale

Fri May 19 10:53:29 1995

800

810

Page
45



maryh@stic

stdin

NeWSprinter20

Fri May 19 11:01:15 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5

11. R44469 Eleusine indica alpha-1-tubul 451 8 8 4.20 0
 12. R53465 Glutamate-1-semialdehyde amin 441 8 8 4.20 0
 13. R27872 Odorant receptor clone I7. 327 8 8 4.20 0

The list of other best scores is:

Sequence Name	Description	Length	Score	Init. Opt.	Sig.	Frame
**** 2 standard deviations above mean ****						
14. R38905	Diuretic hormone binding pept	6	5	5	2.40	0
15. R51613	Mimotope nonapeptide #19 from	9	5	5	2.40	0
16. R55669	30-36 KD glial growth factor	10	5	5	2.40	0
17. R46862	Cell proliferation inhibitor.	10	5	5	2.40	0
18. R43629	Peptide derived from insulin-	10	5	5	2.40	0
19. R36885	Insulin-like growth factor-II	10	5	5	2.40	0
20. R28319	Novel peptide GGF-I 13.	10	5	5	2.40	0
21. R28626	Alpha-1 collagen chain derive	10	5	5	2.40	0
22. P60730	Block unit.	10	5	5	2.40	0
23. R37523	22(D) 3-(2'naphthyl)-alanine29A	11	5	5	2.40	0
24. R37522	22(D) 3-(2'naphthyl)-alanine29A	11	5	5	2.40	0
25. R37521	27Tyr. 28Pro. 29Arg. 32Thr. 3	11	5	5	2.40	0
26. R37002	22(D) 3-(2'Naphthyl)-ala 29Ala.	11	5	5	2.40	0
27. R34658	Heparinase tryptic fragment t	11	5	5	2.40	0
28. R27985	Tryptic peptide fragment #3.	11	5	5	2.40	0
29. R49866	Sequence of tryptic digest pe	12	5	5	2.40	0
30. R12491	Lys (342)-alpha-1-antitrypsin d	12	5	5	2.40	0
31. R11088	Human beta-lipotrophic derived	12	5	5	2.40	0
32. R54967	Sorhl grass pollen allergen T	13	5	5	2.40	0
33. P81289	Atrial natriuretic polypeptid	13	5	5	2.40	0
34. R54139	CH255 light chain variable r	14	5	8	2.40	0
35. R36559	Thrombin inhibitor protein pa	14	5	5	2.40	0
36. R34532	Beta-hexosaminidase active pe	14	5	5	2.40	0
37. R32168	Soluble Kit Ligand M1.	14	5	5	2.40	0
38. R22078	Peptide for diagnosis of dise	14	5	5	2.40	0
39. R58349	Porcine parvovirus peptide 13	15	5	5	2.40	0
40. R58347	Porcine parvovirus peptide 8L	15	5	5	2.40	0
41. R58346	Porcine parvovirus peptide 7L	15	5	5	2.40	0
42. R33260	Sj23-like protein fragment.	15	5	5	2.40	0
43. P50768	Synthetic interleukin-2 antig	15	5	5	2.40	0
44. R58348	Porcine parvovirus peptide 10	16	5	5	2.40	0
45. R45248	HCV NS5 region (2422-2437).	16	5	5	2.40	0

1. US-08-121-713B-12 (1-8)
 R12609 Versican.

ID R12609 standard; Protein; 2409 AA.
 AC R12609;
 DT 11-SEP-1991 (first entry)
 DE Versican.
 KW Fibroblast proteoglycan; hyaluronic acid binding;
 KW tissue reconstruction.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..20
 FT /label= signal peptide
 FT Protein 21..2409
 FT /label= versican
 FT Modified -site 34..35

/label= glycosaminoglycan attachment site
 Modified -site 37..38
 /label= glycosaminoglycan attachment site
 Modified -site 507..508
 /label= glycosaminoglycan attachment site
 Modified -site 525..526
 /label= glycosaminoglycan attachment site
 Modified -site 561..562
 /label= glycosaminoglycan attachment site
 Modified -site 644..646
 /label= glycosaminoglycan attachment site
 Modified -site 654..656
 /label= glycosaminoglycan attachment site
 Modified -site 849..851
 /label= glycosaminoglycan attachment site
 Modified -site 948..949
 /label= glycosaminoglycan attachment site
 Modified -site 971..973
 /label= glycosaminoglycan attachment site
 Modified -site 1023..1025
 /label= glycosaminoglycan attachment site
 Modified -site 1047..1048
 /label= glycosaminoglycan attachment site
 Modified -site 1094..1095
 /label= glycosaminoglycan attachment site
 Modified -site 1267..1268
 /label= glycosaminoglycan attachment site
 Modified -site 1270..1272
 /label= glycosaminoglycan attachment site
 Modified -site 1336..1339
 /label= glycosaminoglycan attachment site
 Modified -site 1441..1442
 /label= glycosaminoglycan attachment site
 Modified -site 1444..1446
 /label= glycosaminoglycan attachment site
 Modified -site 1531..1532
 /label= glycosaminoglycan attachment site
 Modified -site 1651..1652
 /label= glycosaminoglycan attachment site
 Modified -site 1780..1781
 /label= glycosaminoglycan attachment site
 Modified -site 1814..1816
 /label= glycosaminoglycan attachment site
 Modified -site 1846..1847
 /label= glycosaminoglycan attachment site
 Modified -site 1851..1852
 /label= glycosaminoglycan attachment site
 Modified -site 1868..1869
 /label= glycosaminoglycan attachment site
 Modified -site 1954..1955
 /label= glycosaminoglycan attachment site
 Modified -site 2133..2134
 /label= glycosaminoglycan attachment site
 Modified -site 2255..2256
 /label= glycosaminoglycan attachment site
 Modified -site 57
 /label= N-glycosylation site
 Modified -site 330
 /label= N-glycosylation site
 Modified -site 411

FT /label= N-glycosylation site
 FT Modified -site 455
 FT /label= N-glycosylation site
 FT Modified -site 481
 FT /label= N-glycosylation site
 FT Modified -site 676
 FT /label= N-glycosylation site
 FT Modified -site 911
 FT /label= N-glycosylation site
 FT Modified -site 1192
 FT /label= N-glycosylation site
 FT Modified -site 1285
 FT /label= N-glycosylation site
 FT Modified -site 1293
 FT /label= N-glycosylation site
 FT Modified -site 1373
 FT /label= N-glycosylation site
 FT Modified -site 1398
 FT /label= N-glycosylation site
 FT Modified -site 1405
 FT /label= N-glycosylation site
 FT Modified -site 1509
 FT /label= N-glycosylation site
 FT Modified -site 1641
 FT /label= N-glycosylation site
 FT Modified -site 1947
 FT /label= N-glycosylation site
 FT Modified -site 2080
 FT /label= N-glycosylation site
 FT Modified -site 2367
 FT /label= N-glycosylation site
 FT Modified -site 2382
 FT /label= N-glycosylation site
 FT Modified -site 2392
 FT /label= N-glycosylation site
 FT Modified -site 416..419
 FT /label= O-glycosylation site
 FT Modified -site 590..592
 FT /label= O-glycosylation site
 FT Modified -site 646..648
 FT /label= O-glycosylation site
 FT Modified -site 729..732
 FT /label= O-glycosylation site
 FT Modified -site 755..757
 FT /label= O-glycosylation site
 FT Modified -site 859..861
 FT /label= O-glycosylation site
 FT Modified -site 1008..1010
 FT /label= O-glycosylation site
 FT Modified -site 1018..1020
 FT /label= O-glycosylation site
 FT Modified -site 1032..1034
 FT /label= O-glycosylation site
 FT Modified -site 1127..1129
 FT /label= O-glycosylation site
 FT Modified -site 1205..1208
 FT /label= O-glycosylation site
 FT Modified -site 1341..1344
 FT /label= O-glycosylation site
 FT Modified -site 1399..1401

FT /label= O-glycosylation site
 FT Modified -site 1407..1412
 FT /label= O-glycosylation site
 FT Modified -site 1468..1471
 FT /label= O-glycosylation site
 FT Modified -site 1735..1737
 FT /label= O-glycosylation site
 FT Modified -site 1990..1992
 FT /label= O-glycosylation site
 FT Modified -site 2018..2020
 FT /label= O-glycosylation site
 FT Region 400..408
 FT /label= unusual cluster of Glu residues
 PN WO9108230-A.
 PD 13-JUN-1991.
 PF 27-NOV-1990; U06897.
 PR 27-NOV-1989; US-441179.
 PA (JOLLI-) LA JOLLA CANCER RES.
 PI Ruoslahti E;
 DR WPI; 91-193156/26.
 PT Versican proteoglycan and nucleic acid - having hyaluronic acid
 PT binding region and used in e.g.. tissue reconstruction
 PS Claim 1; Fig 1; 27pp; English.
 CC The new fibroblast proteoglycan comprises a signal peptide,
 CC N-terminal hyaluronic acid domain, C-terminal EGF-like domain,
 CC lectin-like sequence and complement regulatory protein-like domain.
 CC Recombinant protein can be used to determine the presence of hyal-
 CC uronic acid, and as a vehicle to bring other molecules into contact
 CC with hyaluronic acid. Antibodies can be prepared against the N-
 CC terminal or glycosaminoglycan attachment domain. The protein is
 CC rich in glutamic acid and is highly negatively charged with a pI of
 CC 4.2. There are a total of 34 cysteine residues located exclusively
 CC in the N- and C-terminal domains. On the C-terminal side of the
 CC link protein-like sequence is a 200 AA sequence contg. two
 CC cysteines and an unusual cluster of glutamic acid residues. The
 CC glycosaminoglycan attachment sites have a consensus sequence of
 CC E/DGSGE/D and the site of attachment is the Ser residue.
 SQ Sequence 2409 AA;
 SQ 146A; 79 R; 73 N; 141D; 0 B; 35 C; 258E; 0 Z; 133G; 57 H;
 SQ 114I; 139L; 96 K; 37 M; 95 F; 155F; 253S; 259T; 24 W; 62 Y; 159V;
 Initial Score = 8 Optimized Score = 8 Significance = 4.20
 Residue Identity = 37% Matches = 3 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0
 X X
 XSXTXAXX
 | |
 SVTSTTLEILSDTGAEGPTVAPLPFST
 1340 1350 1360
 2. US-08-121-713B-12 (1-8)
 R42839 Urea amidolyase.
 ID R42839 standard; Protein; 1835 AA.
 AC R42839;
 DT 10-MAY-1994 (first entry)
 DE Urea amidolyase.
 KW Urea amidolyase; URL; yeast; recombinant plasmid.

OS Saccharomyces cerevisiae.
PN J05244959-A.
PD 24-SEP-1993.
PF 05-MAR-1992; 084531.
PR 05-MAR-1992; JP-084531.
PA (TOYM) TOYOCO KK.
DR WPI; 93-338925/43.
DR N-PSDB; Q49460.
PT DNA having the genetic information of urea amidolyase originated from Saccharomyces yeast - can be used to prepare high purity urea amidolyase by culturing the transformant comprising the DNA Claim 2; Page 10-17; 1pp; Japanese.
PS This sequence represents a protein which has urea amidolyase (URL)-CC activity and is derived from yeast. The DNA encoding this protein CC may be used within a recombinant plasmid for the production of highly CC pure URL.
SQ Sequence 1835 AA;
SQ 127A; 74 R; 82 N; 97 D; 0 B; 31 C; 60 Q; 126E; 0 Z; 145G; 23 H;
SQ 127I; 156L; 117K; 26 M; 72 F; 102P; 144S; 95 T; 22 W; 65 Y; 144V;

Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X X
XSXTXAXX
| | |
PVLFNSAVENLSRTGNAVIEIDFEPILE
280 290 300

3. US-08-121-713B-12 (1-8)
R41732 High molecular weight protein 4 (HMW4).

ID R41732 standard; Protein; 1529 AA.
AC R41732;
DT 26-APR-1994 (first entry)
DE High molecular weight protein 4 (HMW4).
KW HMW; high molecular weight protein; virus; vaccine; influenza;
OS epitope; immunity; haemophilus influenzae.
PN WO9319090-A.
PD 30-SEP-1993.
PR 16-MAR-1993; U02166.
PR 16-MAR-1992; GB-005704.
PA (BARE/) BARENKAMP S J.
PI (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
DR Barenkamp SJ;
DR WPI; 93-320683/40.
DR N-PSDB; Q49511.
PT High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
PS Claim 6; Figure 10; 100pp; English.
CC The isolation and purification of the high molecular weight protein CC enables the identification of the major protective epitopes of the CC protein by conventional epitope mapping. These epitopes can then be CC synthesised using standard techniques and incorporated into fully CC synthetic or recombinant vaccines.
SQ Sequence 1529 AA;
SQ 117A; 30 R; 176N; 80 D; 0 B; 2 C; 44 Q; 65 E; 0 Z; 168G; 13 H;

SQ 139I; 88 L; 110K; 6 M; 35 F; 12 P; 148S; 183T; 6 W; 10 Y; 97 V;
Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X X
XSXTXAXX
| | |
TLGNISVEGNLSLTGRANIVGNLSIAE
810 820 X 830

4. US-08-121-713B-12 (1-8)
P81854 Sequence encoded by LAV EL I POL gene

ID P81854 standard; protein; 1022 AA.

AC P81854;
DT 16-DEC-1990 (first entry)
DE Sequence encoded by LAV EL I POL gene
KW HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.
OS Lymphadenopathy associated virus EL I.
PN W08707906-A.
PD 30-DEC-1987.
PF 22-JUN-1987; E00326.
PR 23-JUN-1986; EP-401380.
PA (INSP) Inst Pasteur.
PI Alizon M, Sonigo P, Wain-Hobson S, Montagnier L;
DR WPI; 88-014396/02.
DR N-PSDB; N80436.

PT New variants of lymphadenopathy associated virus (LAV) -
PT used for prodn. of DNA, antigens and antibodies used in
PT diagnosis of AIDS and pre-AIDS
PS Claim 8; Fig 7A-7J; 72pp; English.

CC LAV EL I (N80436) and LAV MA L (N80437) were isolated from the peripheral
CC blood lymphocytes of patients. Different AIDS virus isolates concerned
CC are designated by 3 letters of the patients name. Stable probes including
CC the DNA sequences can be used for detection of the new LAV viruses or
CC related viruses or DNA proviruses in e.g. biological samples. Proteins
CC or peptides can be used for detection of antibodies induced in vivo and
CC present in biological fluids. The DNA can also be used for the expression
CC of LAV viral antigens for the prodn. of a vaccine against LAV. The
CC polypeptides can also be used for the prodn. of antibodies for the
CC detection of proteins related to the LAV viruses, partic. for diagnosis
CC of AIDS or pre-AIDS.

SQ Sequence 1022 AA;
SQ 61 A; 50 R; 37 N; 50 D; 0 B; 10 C; 65 Q; 76 E; 0 Z; 74 G; 19 H;
SQ 80 I; 80 L; 94 K; 14 M; 30 F; 58 P; 39 S; 56 T; 30 W; 30 Y; 69 V;

Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X X
XSXTXAXX
| | |
ELRWGGRDNFLSKTGAERQCTVSFNFPQ
30 40 X 50

5. US-08-121-713B-12 (1-8)
R08060 HIV-1 pol protein of HIVMAL.

ID R08060 standard; protein; 1003 AA.
AC R08060;
DT 18-JAN-1991 (first entry)
DE HIV-1 pol protein of HIVMAL.
KW HIV diagnosis; HIV-pol; vaccine; HIVMAL;
KW protein processing; reverse transcriptase; RNase; integrase.
OS Human immunodeficiency virus - 1.
PN W09010230-A.
PD 07-SEP-1990.
PF 23-FEB-1990; CA0062.
PR 18-APR-1989; GB-008725.
PA (UYOT-) UNIV OF OTTAWA.
PI Kang CY;
DR WPI; 90-290460/38.
PT Improved polypeptide reagent for HIV diagnosis and vaccine -
PT comprises portions of all 4 enzymes encoded by HIV-pol gene
PS Disclosure; Page 11-23; 37pp; English.
CC Several strains of HIV-1 were cloned and the corresponding amino
CC acid sequence derived from the determined DNA sequences.
CC An improved polypeptide reagent comprises portions of all of the
CC 4 enzymes, and is used in a diagnostic test for HIV infection.
CC The peptide is also used in vaccines.
CC See also R08053-63.
SQ Sequence 1003 AA;
SQ 61 A; 35 R; 36 N; 49 D; 0 B; 10 C; 65 Q; 75 E; 0 Z; 75 G; 17 H;
SQ 85 I; 71 L; 97 K; 14 M; 29 F; 55 P; 41 S; 58 T; 30 W; 31 Y; 69 V;
Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| | |
ELRVWGDRTLTGTGAERQGVFSFPQ
40 50

6. US-08-121-713B-12 (1-8)
R08061 HIV-1 pol protein of HIVELI.

ID R08061 standard; protein; 1003 AA.
AC R08061;
DT 18-JAN-1991 (first entry)
DE HIV-1 pol protein of HIVELI.
KW HIV diagnosis; HIV-pol; vaccine; HIVELI;
KW protein processing; reverse transcriptase; RNase; integrase.
OS Human immunodeficiency virus - 1.
PN W09010230-A.
PD 07-SEP-1990.
PF 23-FEB-1990; CA0062.
PR 18-APR-1989; GB-008725.
PA (UYOT-) UNIV OF OTTAWA.
PI Kang CY;
DR WPI; 90-290460/38.
PT Improved polypeptide reagent for HIV diagnosis and vaccine -
PT comprises portions of all 4 enzymes encoded by HIV-pol gene

PS Disclosure; Page 11-23; 37pp; English.
CC Several strains of HIV-1 were cloned and the corresponding amino
CC acid sequence derived from the determined DNA sequences.
CC An improved polypeptide reagent comprises portions of all of the
CC 4 enzymes, and is used in a diagnostic test for HIV infection.
CC The peptide is also used in vaccines.
CC See also R08053-63.
SQ Sequence 1003 AA;
SQ 61 A; 44 R; 38 N; 48 D; 0 B; 10 C; 63 Q; 75 E; 0 Z; 73 G; 17 H;
SQ 81 I; 78 L; 92 K; 15 M; 29 F; 57 P; 39 S; 55 T; 30 W; 30 Y; 68 V;
Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| | |
ELRVWGRDNPLSKTGAERQGVTSFNFPQ
40 50

7. US-08-121-713B-12 (1-8)
P81861 Sequence encoded by LAV MA L POL gene

ID P81861 standard; protein; 1002 AA.
AC P81861;
DT 16-DEC-1990 (first entry)
DE Sequence encoded by LAV MA L POL gene
KW HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.
OS Lymphadenopathy associated virus MA L.
PN W08707906-A.
PD 30-DEC-1987.
PF 22-JUN-1987; E00326.
PR 23-JUN-1986; EP-401380.
PA (INSP) Inst Pasteur.
PI Alizon M, Sonigo P, Wain-Hobson S, Montagnier L;
DR WPI; 88-014396/02.
DR N-PSDB; N80437.
PT New variants of lymphadenopathy associated virus (LAV) -
PT used for prodn. of DNA, antigens and antibodies used in
PT diagnosis of AIDS and pre-AIDS
PS Claim 8; Fig 8A-8I; 72pp; English.
CC LAV EL I (N80436) and LAV MA L (N80437) were isolated from the peripheral
CC blood lymphocytes of patients. Different AIDS virus isolates concerned
CC are designated by 3 letters of the patients name. Stable probes including
CC the DNA sequences can be used for detection of the new LAV viruses or
CC related viruses or DNA proviruses in eg. biological samples. The proteins
CC or peptides can be used for detection of antibodies induced in vivo and
CC present in biological fluids. The DNA can also be used for the expression
CC of LAV viral antigens for the prodn. of a vaccine against LAV. The
CC polypeptides can also be used for the prodn. of antibodies for the
CC detection of proteins related to the LAV viruses, partic. for diagnosis
CC of AIDS or pre-AIDS.
SQ Sequence 1002 AA;
SQ 61 A; 38 R; 36 N; 49 D; 0 B; 10 C; 65 Q; 75 E; 0 Z; 75 G; 17 H;
SQ 84 I; 71 L; 95 K; 13 M; 29 F; 56 P; 40 S; 58 T; 29 W; 31 Y; 70 V;
Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5

Gaps = 0 Conservative Substitutions = 0

 X X
 XSXTXAXX
 | |
ELRVWGDKTLSETGAEROGIVSFSPQ
30 40 X 50

8. US-08-121-713B-12 (1-8)
R09301 Sequence deduced from pol gene of HIV 1-NDK.

ID R09301 standard; protein; 982 AA.
AC R09301;
DE 27-FEB-1991 (first entry)
DE Sequence deduced from pol gene of HIV 1-NDK.
KW Human immunodeficiency virus; AIDS.
OS HIV 1-NDK.
PN W09013630-A.
PD 15-NOV-1990.
PF 02-MAY-1990; F00312.
PR 03-MAY-1989; FR-005914.
PA (INRM) INSERM INST NAT SANTE.
PI Barre-Sinoussi F, Chermann JC, Devaux C, Rey F, Sire J;
PI Spire B;
DR WPI; 90-361470/48.
DR N-PSDB; Q06635.
PT New HIV-NDK retrovirus and protein component - used in vaccines
PT against immuno-deficiency disorders and in raising MAb for
PT retro-virus detection in vivo.
PS Disclosure; Fig 2; 37pp; French.
CC The HIV NDK virus was isolated from peripheral blood lymphocytes of
CC an AIDS patient. A genomic library was prep'd. from DNA extracted
CC from CEM cells infected with the virus. The library was screened
CC with a pB1 probe corresp. to a fragment from HIV 1. The virus is
CC more cytopathic than other strains and is not inhibited by OKT4A.
CC It has been deposited as CNCM I-857. The sequence can be used to
CC express proteins useful for diagnosing the presence of NDK and
CC related viruses and in vaccines against immunodeficiency diseases.
CC See also R09301-5.
SQ Sequence 982 AA;
SQ 58 A; 39 R; 34 N; 47 D; 0 B; 10 C; 62 Q; 77 E; 0 Z; 74 G; 16 H;
SQ 79 I; 75 L; 89 K; 12 M; 31 F; 54 P; 39 S; 58 T; 28 W; 31 Y; 69 V;

Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

 X X
 XSXTXAXX
 | |
ELRVWGDKNPLSETGAEROGIVSFSPQ
30 40 X 50

9. US-08-121-713B-12 (1-8)
R54631 Fragment of mannuronan C-5-epimerase (epimerase 3)

ID R54631 standard; Protein; 872 AA.
AC R54631;

DT 04-JAN-1995 (first entry)
DE Fragment of mannuronan C-5-epimerase (epimerase 3).
KW Mannuronan; epimerase; alginates; immunostimulant;
KW cell immobilisation; Azotobacter vinelandii; guluronic acid;
KW mannuronic acid.
OS Azotobacter vinelandii strain E.
PN W09409124-A.
PD 28-APR-1994.
PF 08-OCT-1993; N00151.
PR 08-OCT-1992; GB-021163.
PA (NOBI-) NOBIPOL NOBIPOLS FORSKNINGSSSTIFTELSE.
PA (FRON-) PRONOVA BIOPOLYMER AS.
PI Ertesvag H, Larsen B, Valla S, Skjak-braek G;
DR WPI; 94-151310/18.
DR N-PSDB; Q63293.
PT New nucleic acid encoding mannuronan C-5-epimerase - and derived
PT vectors and polypeptide, for prodn. of alginate of controlled G
PT or M content, useful e.g. as immunostimulants or for cell
PT immobilisation.
PS Claim 14; Page 42-45; 78pp; English.
CC The epimerase enzymes (See R54628-R54631) can be used for the
CC microbial production of alginates, esp. those of G (guluronic acid)
CC block content 75-98% or those with a specific M (mannuronic acid)/G
CC block content. Inactivation of specific enzyme activities allows
CC prodn. of pure poly-M alginates or those of G block content below
CC 25%. Alginates of high M content are powerful, nontoxic
CC immunostimulants while those of high G content can be used for cell
CC immobilisation.
SQ Sequence 872 AA;
SQ 86 A; 35 R; 55 N; 97 D; 0 B; 0 C; 32 Q; 35 E; 0 Z; 138G; 9 H;
SQ 32 I; 76 L; 9 K; 7 M; 37 F; 12 P; 54 S; 67 T; 3 W; 37 Y; 51 V;

Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

 X X
 XSXTXAXX
 | |
TLEGTAGNDVLSTGTAHELILGLAGNDR
390 X 400 X 410

10. US-08-121-713B-12 (1-8)
R44470 Eleusine indica alpha-1-tubulin.

ID R44470 standard; Protein; 451 AA.
AC R44470;
DT 29-JUN-1994 (first entry)
DE Eleusine indica alpha-1-tubulin.
KW Alpha-1-tubulin; dinitroaniline; phosphorothioamide;
KW herbicide resistance; pesticide resistance; crop improvement;
KW transgenic plant; maize; sugarbeet; ss.
OS Eleusine indica.
FH Key Location/Qualifiers
FT Modified site 239..239
FT /note="Amino acid substitution from Thr"
PN W09324637-A.
PD 09-DEC-1993.
PF 26-MAY-1993; G01085.

PR 26-MAY-1992; GB-011130.
PA (ELLI/) ELLIS R R.
PA (ZENE) ZENECA LTD.
PI Cronin KE, Hussey PJ, Ray JA, Waldin TR;
DR WPI; 93-405828/50.
DR N-PSDB; Q53405.
PT Modified tubulin conferring resistance to anti-tubulin agents,
PT pref. in maize - enables use of di:nitroaniline and
PT phosphoro:thio:amide herbicide during maize growth for weed
PT control
PS Disclosure; Table 6; 53pp; English.
CC Alpha-1-tubulin isotype protein found in the dinitroaniline-
CC susceptible (S) biotype of E. indica. The protein may be modified
CC via amino acid substitution at position 239 from Thr to Ile to
CC resulting in a modified alpha-1-tubulin resistant to dinitroaniline and
CC phosphorothioamide herbicides. The modified protein may be
CC expressed in transgenic maize (Zea mays) or sugarbeet (Beta
CC vulgaris) in order to confer resistance to dinitroaniline and
CC phosphorothioamide. The modified resistant alpha-1-tubulin
CC protein sequence is given in R44469.
SQ Sequence 451 AA;
SQ 33 A; 21 R; 17 N; 29 D; 0 B; 12 C; 14 Q; 35 E; 0 Z; 39 G; 12 H;
SQ 23 I; 35 L; 19 K; 10 M; 20 F; 17 P; 32 S; 23 T; 3 W; 18 Y; 39 V;
Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X X
XSXTXAXX
| | |
GGDDAFNTFFSETGCKHVPRAVFVDL
50 X 60 70

11. US-08-121-713B-12 (1-9)
R44469 Eleusine indica alpha-1-tubulin (modified).
ID R44469 standard; Protein; 451 AA.
AC R44469;
DE 30-AUG-1994 (first entry)
KW Eleusine indica alpha-1-tubulin (modified).
KW Alpha-1-tubulin; dinitroaniline; phosphorothioamide;
KW herbicide resistance; pesticide resistance; crop improvement;
KW transgenic plant; maize; sugarbeet; goosegrass.
OS Eleusine indica.
PN WO9324637-A.
PD 09-DEC-1993.
PF 26-MAY-1993; G01085.
PR 26-MAY-1992; GB-011130.
PA (ELLI/) ELLIS R R.
PA (ZENE) ZENECA LTD.
PI Cronin KE, Ellis JR, Hussey PJ, Ray JA, Waldin TR;
DR WPI; 93-405828/50.
DR N-PSDB; Q53404.
PT Modified tubulin conferring resistance to anti-tubulin agents,
PT pref. in maize - enables use of di:nitroaniline and
PT phosphoro:thio:amide herbicide during maize growth for weed
PT control
PS Disclosure; Table 3; 53pp; English.

CC DNA encoding an alpha-1-tubulin isotype is found in the dinitroaniline-
CC susceptible (S) biotype of E. indica. The protein generated from
CC this sequence may be modified via amino acid substitution to
CC generate a modified alpha-1-tubulin resistant to dinitroaniline and
CC phosphorothioamide herbicides. The modified DNA sequence of this
CC resistant isotype is given in Q53404.
SQ Sequence 451 AA;
SQ 33 A; 21 R; 16 N; 30 D; 0 B; 12 C; 14 Q; 35 E; 0 Z; 39 G; 12 H;
SQ 25 I; 35 L; 20 K; 10 M; 20 F; 18 P; 31 S; 21 T; 3 W; 18 Y; 38 V;
Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X X
XSXTXAXX
| | |
GGDDAFNTFFSETGCKHVPRAVFVDL
50 X 60 70

12. US-08-121-713B-12 (1-8)
R53465 Glutamate-1-semialdehyde aminotransferase.
ID R53465 standard; Protein; 441 AA.
AC R53465;
DE 18-JAN-1995 (first entry)
KW Glutamate-1-semialdehyde aminotransferase.
KW Glutamate-1-semialdehyde aminotransferase; Xanthomonas phaseoli;
KW expression; transformation; fermentation; production.
OS Propionibacterium freudenreichii.
PN J06113861-A.
PD 26-APR-1994.
PF 07-OCT-1992; 268655.
PR 07-OCT-1992; JP-288655.
PA (MAZN) COSMO OIL CO LTD.
PA (COSM-) COSMO SOGO KENKYUSHO KK.
DR WPI; 94-172754/21.
DR N-PSDB; Q63612.
PT DNA fragment coding glutamate-1 semi-aldehyde amino transferase -
PT used for production of alanine and vitamin-B12.
PS Claim 2; Page 5; 9pp; Japanese.
CC The sequence encoding the glutamate-1-semialdehyde aminotransferase
CC (Q63612) was used in the construction of expression vectors which
CC were used, in turn, to transform competent E. coli. Culturing the
CC transformed cells allows the production of large amounts of
CC glutamate-1-semialdehyde aminotransferase having high activity.
SQ Sequence 441 AA;
SQ 68 A; 22 R; 5 N; 31 D; 0 B; 5 C; 10 Q; 19 E; 0 Z; 45 G; 11 H;
SQ 12 I; 41 L; 13 K; 11 M; 19 F; 23 P; 32 S; 19 T; 5 W; 7 Y; 43 V;
Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X X
XSXTXAXX
| | |
IQDEVLTGFRISPTGAWGLOCAKEGWTP
240 X 250 X 260

13. US-08-121-713B-12 (1-8)
R27872 Odorant receptor clone I7.

ID R27872 standard; Protein; 327 AA.
AC R27872;
DE 15-MAR-1993 (first entry)
DE Odorant receptor clone I7.
KW hormone; G-protein; insect; vertebrate; fish; mammal; neurotransmitter;
KW hormone; G-protein; surface receptor; olfactory epithelium; PCR;
KW Sprague-Dawley rat; amplify; primer; polymerase chain reaction;
KW multigene family; ligand binding domain.
OS Ratus ratus.
FH Key Location/Qualifiers
FT Misc difference 35
FT /label= VAL, ALA, ASP, GLY
PN W09217585-A.
PD 15-OCT-1992.
PF 06-APR-1992; U02741.
PF 05-APR-1991; US-681880.
PA (UYCO) UNIV COLUMBIA NEW YORK.
PI Axel R, Buck LB;
DR WPI; 92-366257/44.
DR N-PSDB; Q29860.
PT Nucleic acid encoding an odorant receptor - can be used to
PT control insect populations or for detecting odours e.g. alcohol,
PT explosives, natural gas etc.
PS Claim 41; Fig 14; 195pp; English.
CC The sequences given in R27867-89 are encoded by odorant receptor
CC clones derived from an insect, a vertebrate, a fish or a mammal.
CC These clones form a family of neurotransmitters and hormone receptors
CC which transduce intracellular signals by activation of specific G-
CC proteins. Each of these receptors is a member of a superfamily of
CC surface receptors which traverse the membrane seven times. These
CC clones are only expressed in the olfactory epithelium. These clones
CC were isolated using probes derived from RNA prepared from the
CC olfactory epithelium of Sprague-Dawley rats. Isolated cDNA's were
CC amplified using primers which correspond to transmembrane domain 2
CC and 7. PCR products of the appropriate size were isolated and
CC sequenced. The deduced protein sequences of these cDNA's defined a
CC new multigene family which shared sequence and structural properties
CC with the superfamily of neurotransmitter and hormone receptors which
CC traverse the membrane seven times. This novel family, however
CC exhibits features different from any other member of the superfamily
CC identified so far. There is a striking divergence within the third,
CC fourth and fifth transmembrane domains between the olfactory proteins.
CC This divergence in the potential ligand binding domain is consistent
CC with the idea that the family of molecules cloned is capable of
CC associating with a large number of odorant of diverse molecular
CC structure.
SQ Sequence. 327 AA;
SQ 30 A; 15 R; 13 N; 7 D; 0 B; 10 C; 7 Q; 8 E; 0 Z; 18 G; 10 H;
SQ 25 I; 43 L; 10 K; 12 M; 21 F; 15 P; 24 S; 17 T; 2 W; 13 Y; 26 V;
SQ 1 Others;
Initial Score = 8 Optimized Score = 8 Significance = 4.20
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |
VLAIFILLGPLSVTGASVMAITGAVMRI
210 X 220 X 230

14. US-08-121-713B-12 (1-8)
R38905 Diuretic hormone binding peptide #2.

ID R38905 standard; peptide; 6 AA.
AC R38905;
DE 10-NOV-1993 (first entry)
DE Diuretic hormone binding peptide #2.
KW Diuretic hormone; pharmaceuticals; diagnosis; reagent; detection;
KW precursor; drug; activation; modification.
OS Synthetic.
PN J05140187-A.
PD 08-JUN-1993.
PF 12-NOV-1991; 295719.
PF 12-NOV-1991; JP-295719.
PA (HITB) HITACHI CHEM CO LTD.
DR WPI; 93-216808/27.
PT Peptide which can combine with diuretic hormone - is useful as
PT reagent for detecting diuretic hormone and as treatment drug
PS Claim 1; Page 2; 7pp; Japanese.
CC The sequences given in R38904-05 are peptide fragments which bind
CC to diuretic hormone. These peptides may be incorporated into a
CC larger peptide which can be used in pharmaceuticals, diagnostics, or
CC for the preparation of living body substances. It may also be used
CC as a reagent for the detection of diuretic hormone or precursors and
CC as a treating drug for activation or modification of diuretic
CC hormone function.
SQ Sequence 6 AA;
SQ 1 A; 0 R; 0 N; 0 D; 0 B; 0 C; 1 Q; 2 E; 0 Z; 1 G; 0 H;
SQ 0 I; 0 L; 0 K; 0 M; 0 F; 0 P; 0 S; 1 T; 0 W; 0 Y; 0 V;
Initial Score = 5 Optimized Score = 5 Significance = 2.40
Residue Identity = 33% Matches = 2 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |
EQTGAE
X X

15. US-08-121-713B-12 (1-8)
R51613 Mimotope nonapeptide #19 from panel of maximally d

ID R51613 standard; peptide; 9 AA.
AC R51613;
DE 19-OCT-1994 (first entry)
DE Mimotope nonapeptide #19 from panel of maximally diverse mimotopes.
KW mimotope panel; rational drug design; candidate drug;
KW screening assay; hydrophobicity; antibody repertoire.
OS Synthetic.
FH Key Location/Qualifiers
FT Modified_site 9

FT /note= "amidated"
FN US3300425-A.
PD 05-APR-1994.
PF 13-OCT-1987; 108130.
PR 13-OCT-1987; US-108130.
PR 11-OCT-1988; US-255906.
PR 06-DEC-1989; US-447009.
PA (TERR-) TERRAPIN TECHNOLOGIES INC.
PI Kauvar IM;
DR WPI; 94-109390/13.
PT Screening of candidate drugs for binding to receptor - by
PT comparing inverse image antibody profile of drug with
PT mimotype-binding profile of receptor
PS Example 5; Fig 7; 29pp; English.
CC A panel of 24 nonapeptides (R51595-R51618) was designed to show
CC high diversity in hydrophobic moment and hydrophobic index, as well
CC as charge distribution and size. Sixteen of these peptides were
CC tested for ability to bind the murine antibody Mab 33-6,
CC arbitrarily chosen. Results of a dot-blot showed that 3 of the 16
CC peptides tested successfully bound Mab 33-6. Thus a small number of
CC diverse mimotopes is able to contain a suitable peptide for binding
CC to a selected antibody.
SQ Sequence 9 AA;
SQ 2 A; 1 R; 0 N; 1 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 2 G; 0 H;
SQ 1 I; 0 L; 1 K; 0 M; 0 F; 0 P; 0 S; 1 T; 0 W; 0 Y; 0 V;
Initial Score = 5 Optimized Score = 5 Significance = 2.40
Residue Identity = 25% Matches = 2 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |
GKAITGARD
X X

> O <
O| IO IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq12pir.res made by maryh on Fri 19 May 95 10:44:33-PDT.

Query sequence being compared: US-08-121-713B-12 (1-8)
Number of sequences searched: 75511
Number of scores above cutoff: 3810

Results of the initial comparison of US-08-121-713B-12 (1-8) with:
Data bank : FIR 43, all entries

100000-
N - - *
US0000-
M - -
B - -
E - -
R - -

O - -
F10000* - -
S - -
E 5000-
Q - -
U - -
E - -
N - -
C - -
E 1000-
S - -

500-

100-

50-

10-

5-

SCORE 0 |
SIDEV -1 |

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	6
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
---------	------	--------	--------------------

2 4 1.31

Times: CPU 00:01:14.99 Total Elapsed 00:01:16.00

Number of residues: 22468834
Number of sequences searched: 75311
Number of scores above cutoff: 3810

The scores below are sorted by initial score.
Significance is calculated based on initial score.

111 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig. Frame
1. S17653	dynamin beta heavy chain, cili	4466	8	8	4.59 0
2. S28916	dystrophin - mouse	3678	8	8	4.59 0
3. S18268	delta-(l-alpha-aminoadipyl)-L	3649	8	8	4.59 0
4. A60979	proteoglycan 24K core protein	2409	8	8	4.59 0
5. A46082	urea carboxylase (EC 6.3.4.6)	1835	8	8	4.59 0
6. JN0896	crystalline surface layer pro	1645	8	8	4.59 0
7. RRM096	RNA-directed RNA polymerase	1643	8	8	4.59 0
8. JH0675	restriction precursor - chicke	1353	8	8	4.59 0
9. A53824	pore membrane protein POM152	1337	8	8	4.59 0
10. S26650	DNA-binding protein 5 - human	1203	8	8	4.59 0
11. S31301	DNA repair protein RAD5 - yea	1169	8	8	4.59 0
12. A36096	Ca2+-transporting ATPase (EC	1037	8	8	4.59 0
13. S04111	collagen alpha 2(VI) chain pr	1022	8	8	4.59 0
14. S23378	collagen alpha 2(VI) chain lo	1022	8	8	4.59 0
15. GNLJND	pol polyprotein - human immun	1002	8	8	4.59 0
16. S33121	CloX protein - dog (fragment)	975	8	8	4.59 0
17. S23377	collagen alpha 2(VI) chain sh	918	8	8	4.59 0
18. DEEC	alcohol dehydrogenase (EC 1.1	891	8	8	4.59 0
19. S14809	alcohol dehydrogenase (EC 1.1	891	8	8	4.59 0
20. A35797	probable DNA-binding protein	790	8	8	4.59 0
21. A49398	cycloartenol synthase (EC 5.4	759	8	8	4.59 0
22. A31806	collagen alpha 2(VI) chain -	720	8	8	4.59 0
23. BVEUCB	uvrB protein - Escherichia co	673	8	8	4.59 0
24. S32991	hypothetical protein - human	657	8	8	4.59 0
25. A43317	germ cell-less protein gcl -	569	8	8	4.59 0
26. S36363	pyruvate decarboxylase (EC 4.	564	8	8	4.59 0
27. DCBYP	pyruvate decarboxylase (EC 4.	563	8	8	4.59 0
28. A37293	pyruvate decarboxylase (EC 4.	563	8	8	4.59 0
29. S24344	glucose transport protein glu	528	8	8	4.59 0
30. S05319	glucose transport protein, he	523	8	8	4.59 0
31. S06920	glucose transport protein, he	523	8	8	4.59 0
32. A31556	glucose transport protein, he	522	8	8	4.59 0
33. A45666	variable surface glycoprotein	512	8	8	4.59 0
34. S36339	phosphorin II - Volvox carte	484	8	8	4.59 0
35. S30057	dihydroliipoamide dehydrogenas	479	8	8	4.59 0
36. A28914	tubulin alpha chain - Naegler	453	8	8	4.59 0
37. S07459	tubulin alpha-I chain - Plasm	453	8	8	4.59 0
38. S47489	alpha-tubulin - Naegleria gru	453	8	8	4.59 0
39. S16339	tubulin alpha chain - Toxopla	453	8	8	4.59 0
40. A60671	tubulin alpha chain - sea urc	452	8	8	4.59 0
41. S42033	alpha tubulin - common limpet	452	8	8	4.59 0

42. S45070 tubulin 2 alpha chain - commo 452 8 8 4.59 0
43. S11207 tubulin alpha chain - sea urc 452 8 8 4.59 0
44. S37933 aspartate transaminase (EC 2. 451 8 8 4.59 0
45. S15773 tubulin alpha-1 chain - maize 451 8 8 4.59 0

1. US-08-121-713B-12 (1-8)
S17653 dynein beta heavy chain, ciliary - sea urchin (Tri

ENTRY S17653 #type complete
TITLE dynein beta heavy chain, ciliary - sea urchin (Tripneustes gratillia)
CONTAINS dynein ATPase (EC 3.6.1.33)
ORGANISM #formal name Tripneustes gratillia
DATE 04-Dec-1992 #sequence_revision 02-May-1994 #text_change 08-Dec-1994

ACCESSIONS S17653; S24628

REFERENCE S17653 Gibbons, I.R.; Gibbons, B.H.; Mocz, G.; Asai, D.J.
#authors Nature (1991) 352:640-643
#journal Multiple nucleotide-binding sites in the sequence of dynein
#title beta heavy chain.
#cross-references M01D:91326103

#accession S17653
#molecule type mRNA
#residues 1-4466 ##label GIB1
#cross-references EMBL:X59603
##note the nucleotide sequence is not given in this paper
#accession S24628

##molecule type protein
##residues 162-172;1193-1204;3240-3259;3325-3339 ##label GIB2
CLASSIFICATION #superfamily dynein heavy chain, ciliary
KEYWORDS ATP; heterotrimer; hydrolase; microtubule binding
FEATURE

134-161
1852-1859 #region nucleotide-binding motif A (P-loop)\
2133-2140 #region nucleotide-binding motif A (P-loop)\
2460-2467 #region nucleotide-binding motif A (P-loop)\
2805-2812 #region nucleotide-binding motif A (P-loop)\
1858 #binding_site ATP (Lys) #status predicted\
2139 #binding_site ATP (Lys) #status predicted\
2466 #binding_site ATP (Lys) #status predicted\
2811 #binding_site ATP (Lys) #status predicted\
SUMMARY #length 4466 #molecular-weight 51178 #checksum 9337
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |
| |

YKSCGNIYKGLSOTGAWGCFDEFNRISV
1890 X 1900 1910

2. US-08-121-713B-12 (1-8)
S28916 dystrophin - mouse

ENTRY S28916 #type complete
TITLE dystrophin - mouse
ORGANISM #formal name Mus musculus #common name house mouse
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change 22-Nov-1993

ACCESSIONS S28916
REFERENCE #authors Bies, R.D.; Phelps, S.F.; Cortez, M.D.; Roberts, R.; Caskey, C.T.; Chamberlain, J.S.
#journal Nucleic Acids Res. (1992) 20:1725-1731
#title Human and murine dystrophin mRNA transcripts are differentially expressed during skeletal muscle, heart, and brain development.

#accession S28916
##status preliminary
##residues 1-3678 #label BIE
##cross-references EMBL:M68839

SUMMARY #length 3678 #molecular-weight 42581.5 #checksum 2460
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |

RTKQDRAPLSTTGASASQTVLTQTS
2420 X 2430 2440

3. US-08-121-713B-12 (1-8)
S18268 delta-(L-alpha-aminoacidipyl)-L-cysteiny1-D-valine s

ENTRY S18268 #type complete
TITLE delta-(L-alpha-aminoacidipyl)-L-cysteiny1-D-valine synthetase - Streptomyces lactamdurans
ORGANISM #formal name Streptomyces lactamdurans
DATE 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 14-Sep-1994

ACCESSIONS S18268; S15283
REFERENCE #authors Martin, J.F.
#submission submitted to the EMBL Data Library, January 1991
#accession S18268
#molecule_type DNA
##residues 1-3649 #label MAR
##cross-references EMBL:X57310

REFERENCE S15283
#authors Coque, J.J.R.; Martin, J.F.; Calzada, J.G.; Liras, P.
#journal Mol. Microbiol. (1991) 5:1125-1133
#title The cephamycin biosynthetic genes pcbAB, encoding a large multidomain peptide synthetase, and pcbC of Nocardia lactamdurans are clustered together in an organization different from the same genes in Acremonium chrysogenum and Penicillium chrysogenum.

#cross-references MUID:92065808
#accession S15283
#molecule_type DNA
##residues 224-940;1319-2010;2373-3307 #label COQ

#cross-references EMBL:X57310
##note the nucleotide sequence is not given in this paper
##note the source is designated as Nocardia lactamdurans

GENETICS
#gene pcbAB
CLASSIFICATION #superfamily alpha-aminoacidipyl-cysteiny1-valine synthetase; acetate-CoA ligase homology; gramicidin S synthetase I repeat homology
KEYWORDS cephamycin biosynthesis
FEATURE 236-830 #domain gramicidin S synthetase I repeat homology #label GRS1\
1331-1906 #domain gramicidin S synthetase I repeat homology #label GRS2\
2385-2954 #domain gramicidin S synthetase I repeat homology #label GRS3

SUMMARY #length 3649 #molecular-weight 404084 #checksum 2296
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |

SYPDRIAFMLSDTGAKIVLAGEAHGSR
2490 X 2500 2510

4. US-08-121-713B-12 (1-8)
A60979 proteoglycan 24K core protein precursor - human

ENTRY A60979 #type complete
TITLE proteoglycan 24K core protein precursor - human
ALTERNATE_NAMES fibroblast-chondroitin sulfate proteoglycan core protein; versican core protein
CONTAINS glial hyaluronate binding protein
ORGANISM #formal name Homo sapiens #common name man
DATE 31-Dec-1993 #sequence_revision 03-Feb-1994 #text_change 28-Oct-1994

ACCESSIONS S06014; A60979; A30358; A29348; A45131
REFERENCE S06014
#authors Zimmermann, D.R.; Ruoslahti, E.
#journal EMBO J. (1989) 8:2975-2981
#title Multiple domains of the large fibroblast proteoglycan, versican.

#cross-references MUID:90059882
#accession S06014
#molecule_type mRNA
##residues_ 1-2409 #label ZIM
##cross-references GB:X15998

REFERENCE A60979
#authors Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.
#journal Brain Res. Bull. (1989) 22:67-70
#title Structural similarity of hyaluronate binding proteins in brain and cartilage.

#accession A60979
#molecule_type protein
##residues_ 289-303;171-210 #label BIG

```
REFERENCE
#authors A30358
#journal Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.
#title J. Biol. Chem. (1989) 264:5981-5987
#cross-references Molecule-type protein
#accession A30358
#molecule_type protein
#residues 24-43;40-50;80-87,'D',89-90;80,'E',82-87,'D',89-119;
128-155;167-176,178-180;171-218;229-259,'IR',261-266;
277-283,'G',285-305;328-329,'X',331-340 ##label PER

REFERENCE
#authors A29348
#journal Krusius, T.; Gehlsen, K.R.; Ruoslahti, E.
#title J. Biol. Chem. (1987) 262:13120-13125
#cross-references A fibroblast chondroitin sulfate proteoglycan core protein
#accession A29348 contains lectin-like and growth factor-like sequences.
#molecule_type mRNA
#residues 1725,'V',1727-2409 ##label KRK
#cross-references GB:J02814

REFERENCE
#authors A45131
#journal Perides, G.; Rahemtulla, F.; Lane, W.S.; Asher, R.A.;
#title Bignami, A.
#cross-references Isolation of a large aggregating proteoglycan from human
brain.
#molecule_type protein
#residues 21-22,'X',24-37 ##label PE2
#cross-references NCBI:P118884
#note sequence extracted from NCBI backbone
CLASSIFICATION #superfamily EGF homology; complement factor H repeat
homology; C-type lectin homology
FEATURE
1-20 #domain signal sequence #status experimental #label SIG\
36-348 #domain link protein-like #label LNK\
559-1654 #domain glycosaminoglycan-binding #label GAG\
2106-2137 #domain EGF homology #label EGF\
2144-2175 #domain EGF homology #label EGF2\
2182-2302 #domain C-type lectin homology #label LCH\
2309-2365 #domain complement factor H repeat homology #label FHD
SUMMARY #length 2409 #molecular-weight 265048 #checksum 1473
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X X
X X SXTXAXX
I I I
SVTSTTLIELSDTGAEGFTVAPLPFST
1340 1350 1360
```

5. US-08-121-713B-12 (1-8)
S46082 urea carboxylase (EC 6.3.4.6) - yeast (Saccharomyc

```
ENTRY S46082 #type complete
TITLE urea carboxylase (EC 6.3.4.6) - yeast (Saccharomyc
cerevisiae)
ALTERNATE_NAMES protein YBR1449; protein YBR208C; urea amidolyase
ORGANISM #formal name Saccharomycetes cerevisiae
DATE 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change
08-Dec-1994
ACCESSIONS S46082; S46081; S46080; S31341; S34930; S34033
REFERENCE S45734
#authors Rieger, M.
#submission submitted to the Protein Sequence Database, August 1994
#accession S46082
#molecule_type DNA
#residues 1-893 ##label RIE
#cross-references EMBL:Z36077
REFERENCE S45927
#authors Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
#submission submitted to the Protein Sequence Database, August 1994
#accession S46081
#molecule_type DNA
#residues 873-1835 ##label FEL
#cross-references EMBL:Z36077
REFERENCE S46034
#authors Bussereau, F.; Demolis, N.; Jacquet, M.; Mallet, I.
#submission submitted to the Protein Sequence Database, August 1994
#accession S46080
#molecule_type DNA
#residues 1487-1835 ##label BUS
#cross-references EMBL:Z36077
REFERENCE S31341
#authors Genbauffe, F.S.; Cooper, T.G.
#journal DNA Seq. (1991) 2:19-32
#title The urea amidolyase (DUR1,2) gene of Saccharomycetes
cerevisiae.
#accession S31341
#molecule_type DNA
#residues 1-95,'R',97-255,257-258,'N',259-458,'M',460-829,'K',
831-1394,'E',1396-1835 ##label GEN
#cross-references EMBL:M64926
REFERENCE S34925
#authors Bussereau, F.; Mallet, L.; Gaillon, L.; Jacquet, M.
#journal Yeast (1993) 9:797-806
#title Yeast Sequencing Reports. A 12.8 kb segment, on the right arm
of chromosome II from Saccharomycetes cerevisiae including
part of the DUR1,2 gene, contains five putative new genes.
#accession S34930
#molecule_type DNA
#residues 1487-1835 ##label BU2
#cross-references EMBL:Z21487
#note translation of nucleotide sequence not given
GENETICS
#gene DUR1,2
#map position 2R
KEYWORDS biotin; ligase; membrane protein; P-loop; purine nucleotide
binding
FEATURE
122-129 #region purine nucleotide-binding motif A (P-loop)\
163-179 #domain transmembrane #status predicted #label TM1\
209-225 #domain transmembrane #status predicted #label TM2\
```

```
411-427      #domain transmembrane #status predicted #label TM3\  
1141-1157    #domain transmembrane #status predicted #label TM4\  
1582-1598    #domain transmembrane #status predicted #label TM5\  
1778-1794    #domain transmembrane #status predicted #label TM6\  
128          #binding site ATP/GTP (lys) #status predicted\  
1798         #binding site biotin (lys) (covalent) #status predicted  
SUMMARY      #length 1835 #molecular-weight 201830 #checksum 7842  
SEQUENCE
```

```
Initial Score = 8 Optimized Score = 8 Significance = 4.59  
Residue Identity = 37% Matches = 3 Mismatches = 5  
Gaps = 0 Conservative Substitutions = 0
```

```
          X X  
          XSXTXAXX  
          |||  
PVLFSNAVENLSRTGANVIEIDFEPLLE  
280 290 300
```

```
6. US-08-121-713B-12 (1-8)  
JN0896      crystalline surface layer protein precursor - Rick  
ENTRY       JN0896 #type complete  
TITLE       crystalline surface layer protein precursor - Rickettsia  
ORGANISM     typhi  
DATE         #formal name Rickettsia typhi  
19-May-1994 19-May-1994 #sequence_revision 19-May-1994 #text_change  
ACCESSIONS  JN0896; FN0686  
REFERENCE    Hahn, M.J.; Kim, K.K.; Kim, I.; Chang, W.H.  
#authors     Gene (1993) 133:129-133  
#journal     Cloning and sequence analysis of the gene encoding the  
#title       crystalline surface layer protein of Rickettsia typhi.  
#accession   JN0896  
##molecule_type DNA  
##residues  1-1645 ##label HAH  
#accession   PN0686  
##molecule_type protein  
##residues  1333-1371 ##label HA2
```

```
FEATURE  
1-32      #domain signal sequence #status predicted #label SIG\  
33-1645   #product crystalline surface layer protein #label MAT  
SUMMARY    #length 1645 #molecular-weight 169697 #checksum 9546  
SEQUENCE
```

```
Initial Score = 8 Optimized Score = 8 Significance = 4.59  
Residue Identity = 37% Matches = 3 Mismatches = 5  
Gaps = 0 Conservative Substitutions = 0
```

```
          X X  
          XSXTXAXX  
          |||  
LNSNETADVGGSETGAVSSGDEAIDQVS  
1340 X 1350 1360
```

```
7. US-08-121-713B-12 (1-8)  
RRWGNV      RNA-directed RNA polymerase (EC 2.7.7.48) - narcis
```

```
ENTRY       RRWGNV #type complete  
TITLE       RNA-directed RNA polymerase (EC 2.7.7.48) - narcissus mosaic  
ALTERNATE_NAMES  
ORGANISM     virus  
DATE         RNA nucleotidyltransferase (RNA-directed); RNA replicase  
30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change  
08-Dec-1994
```

```
ACCESSIONS  JT0470
```

```
REFERENCE    JT0470
```

```
#authors     Zuidema, D.; Linthorst, H.J.M.; Huisman, M.J.; Aejes, C.J.;
```

```
#journal     Bol, J.F.
```

```
#title       J. Gen. Virol. (1989) 70:267-276
```

```
#cross-references MUID:89279206
```

```
#accession   JT0470
```

```
##molecule_type genomic RNA
```

```
##residues  1-1643 ##label ZUI
```

```
##cross-references GB:D00405
```

```
CLASSIFICATION #superfamily eggplant mosaic virus RNA-directed RNA
```

```
polymerase
```

```
KEYWORDS     ATP; nucleotidyltransferase; RNA replication; RNA synthesis
```

```
FEATURE  
868-875     #region nucleotide-binding motif A (P-loop)\
```

```
931-936     #region nucleotide-binding motif B\  
874         #binding site ATP (Lys) #status predicted
```

```
SUMMARY      #length 1643 #molecular-weight 186303 #checksum 8014
```

```
SEQUENCE
```

```
Initial Score = 8 Optimized Score = 8 Significance = 4.59
```

```
Residue Identity = 37% Matches = 3 Mismatches = 5
```

```
Gaps = 0 Conservative Substitutions = 0
```

```
          X X  
          XSXTXAXX  
          |||  
CPAAMNANRPLSKTKAMQMLLYCKSVKQ  
290 X 300 X 310
```

```
8. US-08-121-713B-12 (1-8)
```

```
JH0675      restrictin precursor - chicken
```

```
ENTRY       JH0675 #type complete
```

```
TITLE       restrictin precursor - chicken
```

```
ORGANISM     #formal name Gallus gallus #common name chicken
```

```
DATE         09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change
```

```
15-Oct-1994
```

```
ACCESSIONS  JH0675; PS0385
```

```
REFERENCE    JH0675
```

```
#authors     Noerenberg, U.; Wille, H.; Wolff, J.M.; Frank, R.; Rathjen,
```

```
#journal     F.G.
```

```
#title       Neuron (1992) 8:849-863
```

```
#cross-references MUID:92265298
```

```
#contents    The chicken neural extracellular matrix molecule restrictin:
```

```
#accession   similarity with EGF-, fibronectin type III-, and
```

```
JH0675      fibrinogen-like motifs.
```

```
##molecule_type mRNA
```

##residues 1-1353 ##label NOE
##cross-references GB:X64649
#accession PS0385
##molecule type protein
##residues 579-586; 827-840 ##label NOE1
COMMENT This protein is a neural extracellular matrix protein implicated in
neural cell attachment.
CLASSIFICATION #superfamily restrictin; EGF homology; fibrinogen beta/gamma
homology; fibronectin type III repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; extracellular
matrix; glycoprotein; homotrimer; tandem repeat
FEATURE
1-33 #domain signal sequence #label SIG\
34-1353 #product restrictin #status predicted #label MAT\
203-229 #domain EGF homology #label EG1\
234-260 #domain EGF homology #label EG2\
265-291 #domain EGF homology #label EG3\
296-322 #domain EGF homology #label EG4\
323-412 #domain fibronectin type III repeat homology #label FN1\
413-501 #domain fibronectin type III repeat homology #label FN2\
502-591 #domain fibronectin type III repeat homology #label FN3\
592-683 #domain fibronectin type III repeat homology #label FN4\
684-771 #domain fibronectin type III repeat homology #label FN5\
772-860 #domain fibronectin type III repeat homology #label FN6\
861-948 #domain fibronectin type III repeat homology #label FN7\
949-1034 #domain fibronectin type III repeat homology #label FN8\
1035-1122 #domain fibronectin type III repeat homology #label FN9\
1130-1338 #domain fibrinogen beta/gamma homology #label FBG\
1272-1286 #domain calcium-binding #status predicted #label CAL\
53,197,277,391,469,
580,734,790,960,
1031,1041,1256,
1342 #binding site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 1353 #molecular-weight 149278 #checksum 975
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |
VTASTSISLWTKAMGPIDHYRVTF

700 X 710 720

9. US-08-121-713B-12 (1-8)
A53824 pore membrane protein POM152 - yeast (Saccharomyces
ENTRY A53824 #type complete
TITLE pore membrane protein POM152 - yeast (Saccharomyces
cerevisiae)
ORGANISM #formal name Saccharomyces cerevisiae
DATE 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change
07-Oct-1994

ACCESSIONS A53824
REFERENCE A53824
#authors Wozniak, R.W.; Blobel, G.; Rout, M.P.

##journal J. Cell Biol. (1994) 125:31-42
##title POM152 is an integral protein of the pore membrane domain of
the yeast nuclear envelope.
#accession A53824
##status preliminary
##molecule_type DNA
##residues 1-1337 ##label WOZ
##cross-references GB:Z31592
GENETICS
#gene POM152
KEYWORDS glycoprotein; nucleus; repeat; transmembrane protein
SUMMARY #length 1337 #molecular-weight 151651 #checksum 61
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |
CVGQVGLNFELSFTCAPPVYNTKIYKL
630 640 X 650

10. US-08-121-713B-12 (1-8)
S26650 DNA-binding protein 5 - human
ENTRY S26650 #type complete
TITLE DNA-binding protein 5 - human
ORGANISM #formal name Homo sapiens #common name man
DATE 25-Feb-1994; #sequence_revision 25-Feb-1994; #text_change
25-Feb-1994

ACCESSIONS S26650
REFERENCE S26650
#authors Mattioni, T.; Hume, C.R.; Konigorski, S.; Hayes, P.;
Osterweil, Z.; Lee, J.S.
#journal Chromosoma (1992) 101:618-624
#title A cDNA clone for a novel nuclear protein with DNA binding
activity.

#accession S26650
##status preliminary
##residues 1-1203 ##label MAT
##cross-references EMBL:X63071

SUMMARY #length 1203 #molecular-weight 131866 #checksum 8686
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |

VVPERPVTVMVSETPAMSAEPTVLASEP
160 X 170 X 180

11. US-08-121-713B-12 (1-8)
S31301 DNA repair protein RAD5 - yeast (Saccharomyces cer

ENTRY S31301 #type complete
TITLE DNA repair protein RAD5 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES REV2 protein
ORGANISM #formal name Saccharomyces cerevisiae
DATE 28-May-1993 #sequence_revision 28-May-1993 #text_change 02-Jun-1994

ACCESSIONS S31301; S26983
REFERENCE S31301
#authors Johnson, R.E.; Henderson, S.T.; Petes, T.D.; Prakash, S.; Bankmann, M.; Prakash, L.
#journal Mol. Cell. Biol. (1992) 12:3807-3818
#title Saccharomyces cerevisiae RAD5-encoded DNA repair protein contains DNA helicase and zinc-binding sequence motifs and affects the stability of simple repetitive sequences in the genome.

#cross-references MUID:92373048
#accession S31301
#molecule_type DNA
#residues 1-1169 #label JOH
#cross-references EMBL:M96644

REFERENCE S26983
#authors Ahne, F.; Baur, M.; Eckardt-Schupp, F.
#journal Curr. Genet. (1992) 22:277-282
#title The REV2 gene of Saccharomyces cerevisiae: cloning and DNA sequence.

#accession S26983
#molecule_type DNA
#residues 402-477, 'R', 479-634, 'N', 636-845, 'S', 847-897, 'S', 899-972, 'A', 974-1062, 'R' #label AHN
#cross-references EMBL:S46103

GENETICS LISTA:RAD5; REV2
#gene
#map position 12
#map position 12
#words DNA repair; leucine zipper; zinc finger
SUMMARY #length 1169 #molecular-weight 134001 #checksum 8376
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |
EFAKAASDGEASETGANNINPLWKQFKW
460 X 470 480

12. US-08-121-713B-12 (1-8)
A36096 Ca2+-transporting ATPase (EC 3.6.1.38) - fission y

ENTRY A36096 #type complete
TITLE Ca2+-transporting ATPase (EC 3.6.1.38) - fission yeast (Schizosaccharomyces pombe)
ORGANISM #formal name Schizosaccharomyces pombe
DATE 08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change 30-Sep-1993

ACCESSIONS A36096
REFERENCE A36096

#authors Ghislain, M.; Goffeau, A.; Halachmi, D.; Eilam, Y.
#journal J. Biol. Chem. (1990) 265:18400-18407
#title Calcium homeostasis and transport are affected by disruption of cta3, a novel gene encoding CA(2+)-ATPase in Schizosaccharomyces pombe.

#cross-references MUID:91009335
#accession A36096
#status preliminary
#molecule_type DNA
#residues 1-1037 #label GHI
#cross-references GB:J05634

KEYWORDS hydrolase
SUMMARY #length 1037 #molecular-weight 115326 #checksum 8079
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| |
SIIPESLIIVLSITMAGQKNMSKRRVI
330 X 340

13. US-08-121-713B-12 (1-8)
S04111 collagen alpha 2(VI) chain precursor - chicken

ENTRY S04111 #type complete
TITLE collagen alpha 2(VI) chain precursor - chicken
ORGANISM #formal name Gallus gallus #common name chicken
DATE 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 18-Jun-1993

ACCESSIONS S04111
REFERENCE S04111
#authors Koller, E.; Winterhalter, K.H.; Trueb, B.
#journal EMBO J. (1989) 8:1073-1077
#title The globular domains of type VI collagen are related to the collagen-binding domains of cartilage matrix protein and von Willebrand factor.

#cross-references MUID:89305506
#accession S04111
#molecule_type mRNA
#residues 1-1022 #label KOL
#cross-references EMBL:X15041
#note it is uncertain whether Met-1 or Met-8 is the initiator

FEATURE
1-27 #domain signal sequence #status predicted #label SIG
28-1022 #product collagen alpha 2(VI) chain #status predicted #label MAT

SUMMARY #length 1022 #molecular-weight 109176 #checksum 3489
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX

```
1111
SRGSIADPKSETGARVGWVQVSHGTT
650 X 660 670
```

14. US-08-121-713B-12 (1-8)

```
collagen alpha 2(VI) chain long form precursor - c
S23378      S23378      #type complete
collagen alpha 2(VI) chain long form precursor - chicken
#formal name Gallus gallus #common name chicken
22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993
S23378      S23378      #type complete
collagen alpha 2(VI) chain long form precursor - chicken
#formal name Gallus gallus #common name chicken
22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993
S23378      S23378      #type complete
collagen alpha 2(VI) chain long form precursor - chicken
#formal name Gallus gallus #common name chicken
22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993
```

```
X X
XSXTXAXX
1111
```

```
SRGSIADPKSETGARVGWVQVSHGTT
650 X 660 670
```

15. US-08-121-713B-12 (1-8)

```
GNLJND      GNLJND      #type complete
pol polyprotein - human immunodeficiency virus type 1
(isolate NDK)
endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16);
RNA-directed DNA polymerase (EC 2.7.7.49)
#formal name human immunodeficiency virus type 1, HIV-1
host Homo sapiens (man)
30-Sep-1991; #sequence_revision 30-Sep-1991; #text_change
08-Dec-1994
JQ0067
JQ0067
Spire, B.; Sire, J.; Zachar, V.; Rev, F.; Barre-Sinoussi, F.;
Galibert, F.; Hampe, A.; Chermann, J.C.
Gene (1989) 81:275-284
Nucleotide sequence of HIV1-NDK: a highly cytopathic strain
of the human immunodeficiency virus.
#cross-references MUID:90034200
#accession JQ0067
#molecule_type DNA
#residues 1-1002 ##label SPI
##cross-references GB:M27323
```

```
COMMENT      Specific enzymatic cleavages may yield mature proteins including
               protease, reverse transcriptase, and endonuclease. However, exact
               cleavage sites are undetermined.
GENETICS
#gene pol
CLASSIFICATION #superfamily pol polyprotein
KEYWORDS      acquired immune deficiency syndrome; aspartic proteinase;
               endonuclease; nucleotidyltransferase; polyprotein; reverse
               transcriptase
FEATURE
56-154      #product retropepsin #status predicted #label RTP\
80          #active site Asp (shared with dimeric partner) #status
               predicted
SUMMARY      #length 1002 #molecular-weight 113621 #checksum 9917
SEQUENCE
```

```
Initial Score = 8 Optimized Score = 8 Significance = 4.59
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
```

```
X X
XSXTXAXX
1111
ELRVWGGDNPUSLSETGAERQGTVSFSFFQ
30 40 X 50
> O <
O/I/O Intelligenicities
> O <
```

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sql2apt.res made by on Fri 19 May 95 8:56:51-PDT.

Query sequence being compared:US-08-121-713B-12 (1-8)
Number of sequences searched: 43470
Number of scores above cutoff: 3809

Results of the initial comparison of US-08-121-713B-12 (1-8) with:
Data bank : Swiss-prot 31, all entries

```
100000-
-
N
U50000-
M
B
E
R
O
F10000-
S
E 5000-
Q
U
E
N
*
```

C

E

S 1000-

500-

100-

50-

10-

5-

0

SCORE 0

STDEV -2

Similarity matrix

Mismatch penalty

Gap penalty

Gap size penalty

Cutoff score

Randomization group

Initial scores to save

Optimized scores to save

Scores:

Times:

Number of residues:

Number of sequences searched:

PARAMETERS

Unitary

K-tuple

Joining penalty

Window size

Alignments to save

Display context

Standard Deviation

Total Elapsed

00:00:44.00

1.13

Mean

Median

4

CPU

00:00:44.04

15335248

43470

Number of scores above cutoff: 3809

Cut-off raised to 3.

Cut-off raised to 4.

Cut-off raised to 5.

Cut-off raised to 6.

The scores below are sorted by initial score.

Significance is calculated based on initial score.

90 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig.	Frame
1. DYHC_TRIGR	DYNEIN BETA CHAIN, CILIARY.	4466	8	8	4.42	0
2. ACVS_NOCIA	L-(ALPHA-AMINOADIPYL)-L-CYSTE	3649	8	8	4.42	0
3. PGCS_HUMAN	LARGE FIBROBLAST PROTEOLYCAN	2409	8	8	4.42	0
4. DURI_YEAST	UREA AMIDOLYASE (CONTAINS: UR	1835	8	8	4.42	0
5. RREP_NMV	RNA REPLICATION PROTEIN (CONT	1643	8	8	4.42	0
6. SON_HUMAN	SON PROTEIN (SON3).	1523	8	8	4.42	0
7. CDP_HUMAN	CCAAT DISPLACEMENT PROTEIN (C	1505	8	8	4.42	0
8. P152_YEAST	NUCLEAR ENVELOPE PORE MEMBRAN	1337	8	8	4.42	0
9. RAD5_YEAST	DNA REPAIR PROTEIN RAD5.	1169	8	8	4.42	0
10. ATC3_SCHPO	CALCIUM-TRANSPORTING ATPASE 3	1037	8	8	4.42	0
11. CA26_CHICK	COLLAGEN ALPHA 2(VI) CHAIN PR	1022	8	8	4.42	0
12. POL_HVIEL	POLYPROTEIN (PROTEASE (EC 1002	1002	8	8	4.42	0
13. POL_HVIMA	POLYPROTEIN (PROTEASE (EC 1002	1002	8	8	4.42	0
14. POL_HVIND	POLYPROTEIN (PROTEASE (EC 1002	1002	8	8	4.42	0
15. POL_HVIZ2	POLYPROTEIN (PROTEASE (EC 1002	1002	8	8	4.42	0
16. CDP_CANFA	CCAAT DISPLACEMENT PROTEIN (H	975	8	8	4.42	0
17. ADHE_ECOLI	ALCOHOL DEHYDROGENASE (EC 1.1	890	8	8	4.42	0
18. A20_HUMAN	PUTATIVE DNA BINDING PROTEIN	790	8	8	4.42	0
19. CAS1_ARATH	CYCLOUTENOL SYNTHASE (EC 5.4	759	8	8	4.42	0
20. UVRB_ECOLI	EXCINUCLEASE ABC SUBUNIT B.	673	8	8	4.42	0
21. UL49_EBV	HYPOTHETICAL BFRF2 PROTEIN.	591	8	8	4.42	0
22. GCL1_DROME	GERM CELL-LESS PROTEIN.	569	8	8	4.42	0
23. DCP1_KIDMA	PYRUVATE DECARBOXYLASE ISOZYM	564	8	8	4.42	0
24. DCP1_HANUV	PYRUVATE DECARBOXYLASE (EC 4.	564	8	8	4.42	0
25. DCP2_YEAST	PYRUVATE DECARBOXYLASE ISOZYM	563	8	8	4.42	0
26. DCP1_YEAST	PYRUVATE DECARBOXYLASE ISOZYM	562	8	8	4.42	0
27. GTR7_RAT	GLUCOSE TRANSPORTER TYPE 7, H	528	8	8	4.42	0
28. GTR2_MOUSE	GLUCOSE TRANSPORTER TYPE 2, L	523	8	8	4.42	0
29. GTR2_RAT	GLUCOSE TRANSPORTER TYPE 2, L	522	8	8	4.42	0
30. DLDH_TRYBB	DIHYDROLIPOAMIDE DEHYDROGENAS	479	8	8	4.42	0
31. TBA_NAERG	TUBULIN ALPHA CHAIN.	453	8	8	4.42	0
32. TBA_PLAFK	TUBULIN ALPHA CHAIN.	453	8	8	4.42	0
33. TBA_TOXGO	TUBULIN ALPHA CHAIN.	453	8	8	4.42	0
34. TBA1_PARLI	TUBULIN ALPHA-1 CHAIN.	452	8	8	4.42	0
35. TBA2_MOUSE	TUBULIN ALPHA-2 CHAIN.	451	8	8	4.42	0
36. TBA1_VOLCA	TUBULIN ALPHA-1 CHAIN.	451	8	8	4.42	0
37. TBA1_ANEPH	TUBULIN ALPHA-1 CHAIN.	451	8	8	4.42	0
38. TBA1_CHLRE	TUBULIN ALPHA-1 CHAIN.	451	8	8	4.42	0
39. TBA1_CRIGR	TUBULIN ALPHA-1 CHAIN.	451	8	8	4.42	0
40. TBA1_HUMAN	TUBULIN ALPHA-1 CHAIN.	451	8	8	4.42	0
41. TBA1_MAIZE	TUBULIN ALPHA-1 CHAIN.	451	8	8	4.42	0
42. TBA1_TRYBR	TUBULIN ALPHA-1 CHAIN.	451	8	8	4.42	0
43. TBA_TORMA	TUBULIN ALPHA CHAIN.	451	8	8	4.42	0
44. TBA_PIG	TUBULIN ALPHA CHAIN.	451	8	8	4.42	0

45. TBA_EUGGR TUBULIN ALPHA CHAIN. 451 8 8 4.42 0

1. US-08-121-713B-12 (1-8)

DYHC_TRIGR DYNEIN BETA CHAIN, CILIARY.

ID DYHC_TRIGR STANDARD; PRT; 4466 AA.

AC P23096;

DT 01-NOV-1991 (REL. 20, CREATED)

DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)

DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)

DE DYNEIN BETA CHAIN, CILIARY.

OS TRIPNEUSTES GRATILIA (HAWAIIAN SEA URCHIN).

OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;

CN EUECHINOIDEA.

RN [1]

RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RC TISSUE-BLASTULA EMBRYO;

RM 91326103

RA GIBBONS I.R.; GIBBONS B.H.; MOCZ G.; ASAI D.J.;

RL NATURE 352:640-643(1991).

RN [2]

RP SEQUENCE FROM N.A.

RM 92020893

RA GIBBONS I.R.; ASAI D.J.; CHING N.S.; DOLECKI G.J.; MOCZ G.;

RA PHILLIPSON C.A.; REN H.; TANG W.Y.; GIBBONS B.H.;

RL PROC. NATL. ACAD. SCI. U.S.A. 88:8563-8567(1991).

CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND

CC FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.

CC DYNEIN HAS ATPASE ACTIVITY.

CC -1- SUBUNIT: CONSIST OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA),

CC THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.

CC -1- SIMILARITY: TO OTHER DYNEIN HEAVY CHAINS.

DR EMBL; X59603; TGDYH2.

PIR; S17653; S17653.

KW MOTOR PROTEIN; MICROTUBULES; DYNEIN; ATP-BINDING; FLAGELLA;

KW HEPTAD REPEAT PATTERN.

FT NP_BIND 154 161 ATP (POTENTIAL).

FT NP_BIND 1852 1859 ATP (POTENTIAL).

FT NP_BIND 2133 2140 ATP (POTENTIAL).

FT NP_BIND 2460 2467 ATP (POTENTIAL).

FT NP_BIND 2805 2812 ATP (POTENTIAL).

FT VARIANT 611 615 MISSING.

FT VARIANT 3356 3358 LPG -> LITGNFFCCFWTAG.

SQ SEQUENCE 4466 AA; 511771 MW; 20102986 CN;

Initial Score = 8 Optimized Score = 8 Significance = 4.42

Residue Identity = 37% Matches = 3 Mismatches = 5

Gaps = 0 Conservative Substitutions = 0

X X X

XSXTYAXX

I I I

YKSCNIYKGLSQTGCAMGCFDEFNRISV

1890 X 1900 1910

2. US-08-121-713B-12 (1-8)

ACVS_NOCILA L-(ALPHA-AMINOADIPYL)-L-CYSTEINYL-D-VALINE SYNTHET

ID ACVS_NOCILA STANDARD; PRT; 3649 AA.
AC P27743;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE L-(ALPHA-AMINOADIPYL)-L-CYSTEINYL-D-VALINE SYNTHETASE (EC 6.-.-.-.)
(ACV SYNTHETASE) (ACVS).
DE PCBAB.
GN NOCARDIA LACTAMIDURANS.
OS PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; NOCARDIOFORM.
OC [1]
RN SEQUENCE FROM N.A.
RP STRAIN=VAR LC 411;
RC 92065808
RA COQUE J.J.R.; MARTIN J.F.; CALZADA J.G.; LIRAS P.;

RL MOL. MICROBIOL. 5:1125-1133(1991).

CC -1- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF ACV ARE ACTIVATED

CC AS AMONACYL-ADENYLATES WITH PEPTIDE BONDS FORMED THROUGH THE

CC PARTICIPATION OF AMINO ACID THIOLESTER INTERMEDIATES.

CC -1- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND

CC CEPHALOSPORIN.

CC -1- COFACTOR: CONTAINS THREE COVALENTLY BOUND PHOSPHOPANTHETINES.

CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT

CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

DR EMBL; X57310; NLPCEBAC.

DR PIR; S18268; S18268.

DR PROSITE; PS00455; AMP BINDING.

KW LIGASE; ANTIBIOTIC BIOSYNTHESIS; MULTIFUNCTIONAL ENZYME;

KW REPEAT; PHOSPHOPANTHETINE..

FT REPEAT 401 861 DOMAIN 1.

FT REPEAT 1014 1937 DOMAIN 2.

FT REPEAT 2079 2985 DOMAIN 3.

FT BINDING 820 820 PHOSPHOPANTHETINE (BY SIMILARITY).

FT BINDING 1896 1896 PHOSPHOPANTHETINE (BY SIMILARITY).

FT BINDING 2944 2944 PHOSPHOPANTHETINE (BY SIMILARITY).

SQ SEQUENCE 3649 AA; 404079 MW; 20196076 CN;

Initial Score = 8 Optimized Score = 8 Significance = 4.42

Residue Identity = 37% Matches = 3 Mismatches = 5

Gaps = 0 Conservative Substitutions = 0

X X X

XSXTYAXX

I I I

SYPDRIAFWLSDTGAKLVACEAHCSR

2490 X 2500 2510

3. US-08-121-713B-12 (1-8)

PGCS_HUMAN LARGE FIBROBLAST PROTEOGLYCAN PRECURSOR (VERSICAN)

ID PGCS_HUMAN STANDARD; PRT; 2409 AA.

AC P13611;

DT 01-JAN-1990 (REL. 13, CREATED)

DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)

DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)

DE LARGE FIBROBLAST PROTEOGLYCAN PRECURSOR (VERSICAN) (CHONDROITIN

DE SULFATE PROTEOGLYCAN CORE PROTEIN 2).

GN CSPG2.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RM 90059882
 RA ZIMMERMAN D.R., RUOSLAHTI E.;
 RL EMBO J. 8:2975-2981(1989).
 RN [2]
 RN SEQUENCE OF 1722-2409 FROM N.A.
 RP TISSUE=LUNG FIBROBLAST;
 RM 88007514
 RA KRUSIUS T., GEHLEN K.R., RUOSLAHTI E.;
 RL J. BIOL. CHEM. 262:13120-13125(1987).
 CC -1- SIMILARITY: MAY PLAY A ROLE IN INTERCELLULAR SIGNALING.
 CC -1- SIMILARITY: IN THE N-TERMINAL PART THIS PROTEIN IS SIMILAR TO
 CC A GLIAL HYALURONATE-BINDING PROTEIN. IN THE MIDDLE OF THE PROTEIN
 CC THERE ARE GLUCOSAMINOGLUCAN ATTACHMENT SITES. THE C-TERMINAL PART
 CC CONTAINS EPIDERMAL GROWTH FACTOR (EGF)-LIKE REPEATS AND A LECTIN-
 CC LIKE DOMAIN. VERSCAN CONTAINS A DOMAIN STRONGLY HOMOLOGOUS WITH
 CC THE C-TERMINAL CORE PROTEIN DOMAIN OF THE RAT AND CHICKEN LARGE
 CC AGGREGATING CARTILAGE PROTEOGLYCAN. IT ALSO CONTAINS SEQUENCES
 CC HOMOLOGOUS WITH CHICKEN AND MAMMALIAN HEPATIC AND LUNG LECTINS.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE REPEATS.
 CC -1- SIMILARITY: CONTAINS A C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
 DR EMBL; J02814; HSPG.
 DR PIR; S06014; S06014.
 DR PIR; A29348; A29348.
 DR HSP; P01132; IEGF.
 DR MIM; 118661; 11TH EDITION.
 DR PROSITE; PS00022; EGF.
 DR PROSITE; PS00615; C TYPE LECTIN.
 DR GLYCOPROTEIN; PROTEOGLYCAN; LECTIN; EXTRACELLULAR MATRIX; SUSHI;
 KW SIGNAL; REPEAT; EGF-LIKE DOMAIN.
 FT SIGNAL 1 20 PROBABLE.
 FT CHAIN 21 2409 VERSCAN.
 FT DOMAIN 36 348
 FT DOMAIN 559 1654
 FT DOMAIN 2103 2178
 FT DOMAIN 2179 2305
 FT REPEAT 2308 2366
 FT DISULFID 2309 2352
 FT DISULFID 2338 2365
 FT CARBOHYD 57 57
 FT CARBOHYD 330 330
 FT CARBOHYD 411 411
 FT CARBOHYD 455 455
 FT CARBOHYD 481 481
 FT CARBOHYD 676 676
 FT CARBOHYD 911 911
 FT CARBOHYD 1192 1192
 FT CARBOHYD 1285 1285
 FT CARBOHYD 1293 1293
 FT CARBOHYD 1373 1373
 FT CARBOHYD 1398 1398
 FT CARBOHYD 1405 1405

FT CARBOHYD 1509 1509 POTENTIAL.
 FT CARBOHYD 1641 1641 POTENTIAL.
 FT CARBOHYD 1947 1947 POTENTIAL.
 FT CARBOHYD 2080 2080 POTENTIAL.
 FT CARBOHYD 2382 2382 POTENTIAL.
 FT CARBOHYD 2392 2392 POTENTIAL.
 FT CONFLICT 1722 1726 IKAE -> EFREV (IN REF. 2).
 SQ SEQUENCE 2409 AA; 265048 MW; 21910895 CN;
 Initial Score = 8 Optimized Score = 8 Significance = 4.42
 Residue Identity = 37% Matches = 3 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0
 X X X
 XSXTXAXX
 SVTSTTLIELSDTGAEQTVAPLEFST
 1340 1350 1360
 4. US-08-121-713B-12 (1-8)
 DURI_YEAST UREA AMIDOLYASE (CONTAINS: UREA CARBOXYLASE (EC 6.
 ID DURI YEAST STANDARD; PRT; 1835 AA.
 AC P32528;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE UREA AMIDOLYASE (CONTAINS: UREA CARBOXYLASE (EC 6.3.4.6) /
 DE ALLOPHANATE HYDROLASE (EC 3.5.1.54)).
 GN DURI, 2 OR YBR208C OR YBR1448.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 92199240
 RA GENBAUFFE F.S., COOPER T.G.;
 RL DNA SEQ. 2:19-32(1991).
 RN [2]
 RP SEQUENCE OF 1-893 FROM N.A.
 RC STRAIN-S288C;
 RA RIEGER M.;
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE OF 873-1835 FROM N.A.
 RC STRAIN-S288C;
 RA FELDMANN H., MANNAHAPT G., SCHWARZLOSE C., VETTER I.;
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RP SEQUENCE OF 1487-1835 FROM N.A.
 RC STRAIN-S288C;
 RM 93377417
 RA BUSSEAU F., MALLET L., GAILLON L., JACQUET M.;
 RL YEAST 9:797-806(1993).
 CC -1- FUNCTION:HYDROLYSIS OF UREA TO AMMONIA AND CO2.
 CC -1- CATALYTIC ACTIVITY: ATP + UREA + CO(2) = ADP + ORTHOPHOSPHATE +
 CC UREA-1-CARBOXYLATE.
 CC -1- CATALYTIC ACTIVITY: UREA-1-CARBOXYLATE + H(2)O = 2 CO(2) +
 CC 2 NH(3).
 CC -1- PATHWAY: ALLANTOIN AND ARGININE METABOLISM.

CC -!- SUBUNIT: MONOMER.
CC -!- COFACTOR: BIOTIN.
CC -!- INDUCTION: BY ALLOPHANATE OR ITS NON-METABOLIZED ANALOGUE
CC OXALURATE. RERESSED IN THE PRESENCE OF READILY USED NITROGEN
CC SOURCES.
DR EMBL; M64926; SCUAMD.
DR EMBL; Z36077; SCYBR208C.
DR EMBL; Z21487; SCCHRIIFG.
DR PIR; S46082; S46082.
DR PROSITE; PS00188; BIOTIN.
DR PROSITE; PS00866; CPSASE 1.
DR PROSITE; PS00867; CPSASE 2.
KW LIGASE; HYDROLASE; MULTIFUNCTIONAL ENZYME; ARGININE METABOLISM;
KW BIOTIN; ATP-BINDING.
FT NP BIND 122 129
FT BINDING 1798 1798
FT SIMILAR 1779 1835
FT CONFLICT 96 96
FT CONFLICT 256 258
FT CONFLICT 459 459
FT CONFLICT 830 830
FT CONFLICT 1395 1395
FT CONFLICT 1835 AA; 201831 MW; 16646414 CN;
SQ SEQUENCE 1835 AA; 201831 MW; 16646414 CN;
Initial Score = 8 Optimized Score = 8 Significance = 4.42
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| | |

PVLFSNAVENLSRTGANVIEIDFEPLE
280 290 300

5. US-08-121-713B-12 (1-8)
RRPO_NMV RNA REPLICATION PROTEIN (CONTAINS: RNA-DIRECTED RN

ID RRPO_NMV STANDARD; PRT; 1643 AA.
AC P15035;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE RNA REPLICATION PROTEIN (CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE (EC 2.7.7.48) / PROBABLE HELICASE (186 KD PROTEIN) (ORF 1).
OS NARCISSEUS MOSAIC VIRUS (NMV).
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; POTEXVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RM 89279206
RA ZUIDEMA D., LINTHORST H.J.M., HUISMAN M.J., ASJES C.J., BOL J.F.;
RL J. GEN. VIROL. 70:267-276(1989).
CC -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
CC POSSIBLY FUNCTIONS AS A ATP-BINDING HELICASE.
DR EMBL; D13747; MNCCGAA.
DR EMBL; D00405; NMV.
DR PIR; J0470; RRMGV.
KW ATP-BINDING; HELICASE; RNA REPLICATION; RNA-DIRECTED RNA POLYMERASE.
FT NP BIND 868 875
FT SEQUENCE 1643 AA; 186304 MW; 13928612 CN;

Initial Score = 8 Optimized Score = 8 Significance = 4.42
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| | |
CPAAMNANRPLSKTKAMOMLLYCKSVKQ
290 X 300 X 310

6. US-08-121-713B-12 (1-8)
SON_HUMAN SON PROTEIN (SON3).

ID SON_HUMAN STANDARD; PRT; 1523 AA.
AC P18583;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE SON PROTEIN (SON3).
GN SON.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RM 92049296
RA CHUMAKOV I.M., BERDICHEVSKII F.B., SOKOLOV N.V., REZNIKOV M.V.,
RA PRASOLOV V.S.;
RL MOL. BIOL. (MOSK) 25:731-740(1991).
RN [2]
RP SEQUENCE OF 686-1168 FROM N.A.
RA BERDICHEVSKII F.B., CHUMAKOV I.M., KISELEV L.L.;
RL MOL. BIOL. (MOSK) 22:639-646(1989).
RN [3]
RP ALTERNATIVE SPLICING.
RC TISSUE=PLACENTA;
RA BLISKOVSKII V.V., KIRILLOV A.A., ZACHAR'EV V.M., CHUMAKOV I.M.;
RL MOL. BIOL. (MOSK) 26:547-549(1992).
CC -!- DOMAIN: THE REPEAT REGION CONTAINS FIVE TYPES OF SEQUENCES LARGELY
CC HOMOLOGOUS WITH ONE ANOTHER.
DR EMBL; X63753; HSSONA.
DR EMBL; M36428; HSSON3A.
DR PIR; P00099; P00099.
DR MIM; 182465; 11TH EDITION.
KW REPEAT; DNA-BINDING; ALTERNATIVE SPLICING.
RN [1]
RP DOMAIN 854 1042
FT DOMAIN 854 878
FT DOMAIN 880 905
FT REPEAT 917 923
FT REPEAT 926 944
FT DOMAIN 945 986
FT REPEAT 987 1005
FT DOMAIN 1006 1032
FT REPEAT 1033 1046
FT SIMILAR 689 782
FT SIMILAR 932 995
FT SIMILAR 1046 1105
FT CONFLICT 686 686
TO C-MYC (AA 242-319).
TO GALLIN (AA 2-64).
TO C-MOS (AA 155-213).
A -> R (IN REF. 2).

FT CONFLICT 1083 1083 V -> E (IN REF. 2).
FT CONFLICT 1141 1141 F -> P (IN REF. 2).
SQ SEQUENCE 1523 AA; 168281 MW; 12056646 CN;
Initial Score = 8 Optimized Score = 8 Significance = 4.42
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| | |
VYPERPVTGMSVSETPMSAAPTIVLASEP
270 280 X 290

7. US-08-121-713B-12 (1-8)
CDP_HUMAN CCAAT DISPLACEMENT PROTEIN (CDP).

ID CDP_HUMAN STANDARD; PRT; 1505 AA.
AC P39880;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE CCAAT DISPLACEMENT PROTEIN (CDP).
GN CUTLI.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UMBILICAL VEIN;
RA NEUFELD E.J., SKALNIK D.G., LIEVENS P.M.J., ORKIN S.H.;
RL NATURE GENET. 1:50-55(1992).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: TO DROSOPHILA HOMEBOX PROTEIN CUT.
DR EMBL; M74099; HSCDP.
DR MIM; 116896; 11TH EDITION.
KW HOMEBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN;
KW REPEAT.
FT DNA_BIND 1244 1303 HOMEBOX.
SQ SEQUENCE 1505 AA; 164353 MW; 11395795 CN;
Initial Score = 8 Optimized Score = 8 Significance = 4.42
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| | |
AESYSSQSSLSLTCASRSETPNSPLP
890 X 900 X 910

8. US-08-121-713B-12 (1-8)
P152_YEAST NUCLEAR ENVELOPE PORE MEMBRANE PORTEIN POM152 (P15

ID P152_YEAST STANDARD; PRT; 1337 AA.
AC P39685;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)

DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE NUCLEAR ENVELOPE PORE MEMBRANE PORTEIN POM152 (P150).
GN POM152.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=W303;
RM 94186543
RA WOZNIAK R.W., BLOBEL G., ROUT M.P.;
RL J. CELL BIOL. 125:31-42(1994).
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC MEMBRANE.
CC -!- PTM: THE N-TERMINAL IS BLOCKED.

DR EMBL; Z31592; SCPOM152.
DR LISTA; SC00844; POM152.
KW NUCLEAR PROTEIN; TRANSMEMBRANE; REPEAT; GLYCOPROTEIN.
FT DOMAIN 1 175 PORE SIDE (POTENTIAL).
FT TRANSMEM 176 195 POTENTIAL.
FT DOMAIN 196 1337 CISTERNAL SIDE (POTENTIAL).
FT CARBOHYD 280 280
FT DOMAIN 390 1276 8 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 390 413 1.
FT REPEAT 413 413 2.
FT REPEAT 413 413 3.
FT REPEAT 413 413 4.
FT REPEAT 413 413 5.
FT REPEAT 413 413 6.
FT REPEAT 413 413 7.
FT REPEAT 413 413 8.
SQ SEQUENCE 1337 AA; 151651 MW; 9840385 CN;

Initial Score = 8 Optimized Score = 8 Significance = 4.42
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XSXTXAXX
| | |
CVGQVGLNLFSLTGTGAPYYNTKIYKL
630 640 X 650

9. US-08-121-713B-12 (1-8)
RAD5_YEAST DNA REPAIR PROTEIN RAD5.

ID RAD5_YEAST STANDARD; PRT; 1169 AA.
AC P32849;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE DNA REPAIR PROTEIN RAD5.
GN RAD5 OR REV2 OR SNM2.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RM 92375048
RA JOHNSON R.E., HENDERSON S.T., PETES T.D., PRAKASH S., BANKMANN M.,

RA PRAKASH L.;
RL MOL. CELL. BIOL. 12:3807-3818 (1992).
RN [2]
RP SEQUENCE OF 402-1063 FROM N.A.
RM 93008350
RA AENE F., BAUR M., ECKARDT-SCHUPP F.;
RL CURR. GENET. 22:277-282 (1992).
CC -!- FUNCTION: PROBABLE HELICASE, FUNCTIONS WITH DNA REPAIR PROTEIN
CC RAD18 IN ERROR-FREE POSTREPLICATION DNA REPAIR. INVOLVED IN THE
CC MAINTENANCE OF WILD-TYPE RATES OF INSTABILITY OF SIMPLE
CC REPETITIVE SEQUENCES SUCH AS POLY (GT) REPEATS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC -!- SIMILARITY: TO HELICASES OF THE SNF2/RAD54 FAMILY.
DR EMBL; M96644; SCRAD5A.
DR EMBL; S43248; S43248.
DR EMBL; S46103; S46103.
DR PIR; S31301; S31301.
DR PIR; S26983; S26983.
DR LISTA; SC00924; RAD5.
DR PROSITE; PS00518; ZINC FINGER C3HC4.
KW DNA DAMAGE; DNA REPAIR; NUCLEAR PROTEIN; ZINC-FINGER; DNA-BINDING;
KW HELICASE; ATP-BINDING.
FT DOMAIN 42 60 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 303 315 ARG/LYS-RICH (BASIC).
FT NP BIND 532 539 ATP (POTENTIAL).
FT SITE 681 684 DEGH BOX.
FT ZN FING 914 960 C3HC4-TYPE.
FT CONFLICT 478 478 Q -> R (IN REF. 2).
FT CONFLICT 635 635 T -> N (IN REF. 2).
FT CONFLICT 846 846 G -> S (IN REF. 2).
FT CONFLICT 898 898 R -> S (IN REF. 2).
FT CONFLICT 973 973 V -> A (IN REF. 2).
FT CONFLICT 1063 1063 A -> R (IN REF. 2).
SQ SEQUENCE 1169 AA; 134001 MW; 7092987 CN;
Initial Score = 8 Optimized Score = 8 Significance = 4.42
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X X
XSXTYAXX
| | |
EFKAASDGEASETGANNINPLWKQFKW
460 X 470 480
10. US-08-121-713B-12 (1-8)
ATC3_SCHPO CALCIUM-TRANSPORTING ATPASE 3 (EC 3.6.1.38).
ID ATC3_SCHPO STANDARD; PRT; 1037 AA.
AC P22189;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)
DE CALCIUM-TRANSPORTING ATPASE 3 (EC 3.6.1.38).
GN CTA3.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]

RP SEQUENCE FROM N.A.
RM 91009335
RA GHISLAIN M., GOFFEAU A., HALACHMI D., EILAM Y.;
RL J. BIOL. CHEM. 265:18400-18407 (1990).
CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE
CC HYDROLYSIS OF ATP COUPLED WITH THE TRANSPORT OF THE CALCIUM.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
(EI-E2 ATPASES).
DR EMBL; J05634; SPCA2AT.
DR PIR; A36096; A36096.
DR PROSITE; PS00154; ATPASE EI E2.
KW HYDROLASE; CALCIUM TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
KW MAGNESIUM; ATP-BINDING.
FT MOD RES 368 368 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 1037 AA; 115327 MW; 5734435 CN;
Initial Score = 8 Optimized Score = 8 Significance = 4.42
Residue Identity = 37% Matches = 3 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X X
XSXTYAXX
| | |
SIIPESLIASITMAMGQNMKRRVI
330 X 340
11. US-08-121-713B-12 (1-8)
CA26_CHICK COLLAGEN ALPHA 2 (VI) CHAIN PRECURSOR.
ID CA26_CHICK STANDARD; PRT; 1022 AA.
AC P15988;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE COLLAGEN ALPHA 2 (VI) CHAIN PRECURSOR.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RM 89305506
RA KOLLER E., WINTERHALTER K.H., TRUEB B.;
RL EMBO J. 8:1073-1077 (1989).
RN [2]
RP SEQUENCE OF 8-38 FROM N.A.
RC TISSUE=LIVER;
RM 91187664
RA KOLLER E., HAYMAN A.R., TRUEB B.;
RL NUCLEIC ACIDS RES. 19:485-491 (1991).
CC -!- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1 (VI),
CC ALPHA 2 (VI), AND ALPHA 3 (VI).
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
DR EMBL; X15041; GDCOLGA2.
DR EMBL; X56595; GCOLVIA.
DR PIR; S04111; S04111.

DR PIR; S23378; S23378.
 DR HSP; P19999; 1CIG.
 KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
 KW GLYCOPROTEIN; CELL ADHESION; COLLAGEN; SIGNAL.
 FT SIGNAL 1 27
 FT CHAIN 28 1022 COLLAGEN ALPHA 2(VI) CHAIN.
 FT DOMAIN 28 235 NONHELICAL REGION.
 FT DOMAIN 256 590 TRIPLE-HELICAL REGION.
 FT DOMAIN 591 1022 NONHELICAL REGION.
 FT SITE 514 519 INTERUPTION IN COLLAGENOUS REGION.
 FT DOMAIN 44 168 VWF TYPE A-LIKE DOMAIN.
 FT DOMAIN 613 738 VWF TYPE A-LIKE DOMAIN.
 FT DOMAIN 833 957 VWF TYPE A-LIKE DOMAIN.
 FT CARBOHYD 141 141 POTENTIAL.
 FT CARBOHYD 215 215 POTENTIAL.
 FT CARBOHYD 327 327 POTENTIAL.
 FT CARBOHYD 630 630 POTENTIAL.
 FT CARBOHYD 897 897 POTENTIAL.
 SQ SEQUENCE 1022 AA; 109176 MW; 5032827 CN;
 Initial Score = 8 Optimized Score = 4.42
 Residue Identity = 37% Matches = 3 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X
 XSXTXAXX
 | |
 SRLGSIADPKPKSETGARGVGVQYSHECT
 650 X 660 670

12. US-08-121-713B-12 (1-8)
 POL_HV1EL POL POLYPROTEIN (PROTEASE (EC 3.4.23.-); REVERSE T
 ID POL HV1EL STANDARD; PRT; 1002 AA.
 AC P04589;
 DT 13-AUG-1987 (REL. 05, CREATED)
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE POL POLYPROTEIN (PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE
 DE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).
 GN POL.
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (ELI ISOLATE) (HIV-1).
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
 OC LENTIVIRINAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 86245056
 RA ALIZON M., WAIN-HOBSON S., MONTAGNIER L., SONIGO P.;
 RL CELL 46:63-74(1986).
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE
 CC DETERMINED.
 DR EMBL; X04414; HIVELICG.
 DR HSSP; P03366; 1HRH.
 DR HIV; K03454; POLSELI.
 DR PROSITE; PS00141; ASP PROTEASE.
 DR AIDS; POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;
 KW RNA-DIRECTED DNA POLYMERASE.
 FT CHAIN 56 154
 FT ACT_SITE 80 80 BY SIMILARITY.
 Initial Score = 8 Optimized Score = 4.42
 Residue Identity = 37% Matches = 3 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

13. US-08-121-713B-12 (1-8)
 POL_HV1MA POL POLYPROTEIN (PROTEASE (EC 3.4.23.-); REVERSE T
 ID POL HV1MA STANDARD; PRT; 1002 AA.
 AC P04589;
 DT 13-AUG-1987 (REL. 05, CREATED)
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE POL POLYPROTEIN (PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE
 DE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).
 GN POL.
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (MAL ISOLATE) (HIV-1).
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
 OC LENTIVIRINAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 86245056
 RA ALIZON M., WAIN-HOBSON S., MONTAGNIER L., SONIGO P.;
 RL CELL 46:63-74(1986).
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE
 CC DETERMINED.
 DR EMBL; X04415; HIVMALCG.
 DR HSSP; P03366; 1HRH.
 DR HIV; K03456; POLSMAL.
 DR PROSITE; PS00141; ASP PROTEASE.
 DR AIDS; POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;
 KW RNA-DIRECTED DNA POLYMERASE.
 FT CHAIN 56 134
 FT ACT_SITE 80 80 BY SIMILARITY.
 SQ SEQUENCE 1002 AA; 113537 MW; 5140518 CN;
 Initial Score = 8 Optimized Score = 8 Significance = 4.42
 Residue Identity = 37% Matches = 3 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X
 XSXTXAXX
 | |
 ELRWGDKTLSETGAERGIVSFSPFQ
 30 40 X 50

14. US-08-121-713B-12 (1-8)
 POL_HVIND POL POLYPROTEIN (PROTEASE (EC 3.4.23.-); REVERSE T
 ID POL HVIND STANDARD; PRT; 1002 AA.
 AC P18802;
 Initial Score = 8 Optimized Score = 8 Significance = 4.42
 Residue Identity = 37% Matches = 3 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X
 XSXTXAXX
 | |
 ELRWGDKTLSETGAERGIVSFSPFQ
 30 40 X 50

DT 01-NOV-1990 (REL. 16, CREATED)
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE POL POLYPROTEIN (PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE
 DE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).
 GN POL.
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (NDK ISOLATE) (HIV-1).
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
 OC LENTIVIRINAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 90034200
 RA SPIRE B., SIRE J., ZACHAR V., REY F., BARRE-SINOUSI F., GALIBERT F.,
 RA HAMPE A., CHERMANN J.C.;
 RL GENE 81:275-284(1989).
 CC -!- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE
 CC DETERMINED.
 CC -!- NDK, ISOLATED FROM A ZAIRIAN PATIENT AFFECTED WITH AIDS, AND IS
 CC A HIGHLY CYTOPATHOGENIC STRAIN.
 DR EMBL; M27323; HIVNDK.
 DR PIR; JQ0067; GNIND.
 DR HSP; P03366; 1HRH.
 DR HIV; M27323; POLSNDK.
 DR PROSITE; PS00141; ASP PROTEASE.
 KW AIDS; POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;
 KW RNA-DIRECTED DNA POLYMERASE.
 FT CHAIN 56 154
 FT ACT SITE 80 80 BY SIMILARITY.
 SQ SEQUENCE 1002 AA; 113621 MW; 5174153 CN;

Initial Score = 8 Optimized Score = 8 Significance = 4.42
 Residue Identity = 37% Matches = 3 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

```

X      X
XSXTXAXX
|      |
|      |
ELRVWGDNDPLSETGAERQGTVSFSFPQ
30      40      50

```

15. US-08-121-713B-12 (1-8)
 POL_HV122 POL POLYPROTEIN (PROTEASE (EC 3.4.23.-); REVERSE T

ID POL HV122 STANDARD; PRT; 1002 AA.
 AC P12499;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE POL POLYPROTEIN (PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE
 DE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).
 GN POL.
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (22/CDC-234 ISOLATE) (HIV-1).
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
 OC LENTIVIRINAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA THEODORE T., BUCKLER-WHITE A.;
 RL SUBMITTED (NOV-1988) TO THE HIV DATA BANK.
 CC -!- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE

CC DETERMINED.
 DR EMBL; M22639; REHIV222.
 DR HSP; P03366; 1HRH.
 DR HIV; M22639; POLS2226.
 DR PROSITE; PS00141; ASP PROTEASE.
 KW AIDS; POLYPROTEIN; HYDROLASE; ASPARTYL PROTEASE; ENDONUCLEASE;
 KW RNA-DIRECTED DNA POLYMERASE.
 FT CHAIN 56 154
 FT ACT SITE 81 81 BY SIMILARITY.
 SQ SEQUENCE 1002 AA; 113724 MW; 5178811 CN;

Initial Score = 8 Optimized Score = 8 Significance = 4.42
 Residue Identity = 37% Matches = 3 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

```

X      X
XSXTXAXX
|      |
|      |
ELRVWGDNDPLSETGAERQGTVSFNCPQ
30      40      50

```


maryh@stic

stdin

NeWSprinter20

Fri May 19 10:51:22 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5

8. R44140 Murine FACC encoded by cDNA c 558 5 3.20 0
 9. R51270 Murine FACC encoded by cDNA c 591 5 3.20 0
 10. R30497 N-terminal of LH receptor/FSH 757 5 3.20 0
 11. R05935 Secreted GPIIb subunit of mul 993 5 3.20 0
 12. R44430 eryA region polypeptide modul 3398 5 3.20 0
 **** 2 standard deviations above mean ****
 13. R38556 Sequence of peptide which cor 15 5 2.40 0
 14. R41497 TNF inhibitor peptide IX. 21 5 2.40 0
 15. R25980 Peptide monomer 18. 23 4 5 2.40 0
 16. R21896 Magainin 2 peptide analogue. 23 4 5 2.40 0
 17. R21788 Magainin 2 peptide analogue. 23 4 5 2.40 0
 18. R40010 Scintigraph imaging agent spe 28 4 5 2.40 0
 19. R42547 Leukocyte-binding peptide whi 29 4 5 2.40 0
 20. R42546 Leukocyte-binding peptide whi 30 4 5 2.40 0
 21. R81566 Human insulin acceptor protei 31 4 5 2.40 0
 22. R29178 Astrovirus serotype A1 clone 92 4 5 2.40 0
 23. R25189 21B4 gene clone product pF10 108 4 5 2.40 0
 24. R24082 Truncated TNF-alpha 55KD rece 154 4 5 2.40 0
 25. R24081 Truncated TNF-alpha 55KD rece 158 4 5 2.40 0
 26. R24083 Truncated TNF-alpha 55KD rece 159 4 5 2.40 0
 27. R27496 Native 30 kD TNF inhibitor. 161 4 5 2.40 0
 28. R24080 Truncated TNF-alpha 55KD rece 199 4 5 2.40 0
 29. R80436 Recombinant human glutathioni 200 4 5 2.40 0
 30. R80918 Sequence of human glutathione 201 4 5 2.40 0
 31. R80916 Sequence of bovine glutathione 205 4 5 2.40 0
 32. R82488 Sequence of Mycobacterium tub 205 4 5 2.40 0
 33. R70055 Fes/fps proto-oncogene-relate 211 4 5 2.40 0
 34. R57361 Translation from reading fram 220 4 5 2.40 0
 35. R31229 Prepro-thyrotropin releasing 242 4 5 2.40 0
 36. R22390 Antigen ac-lb. 242 4 5 2.40 0
 37. R82592 Polyhedrin protein. 245 4 5 2.40 0
 38. R32546 HanPV polyhedrin. 246 4 5 2.40 0
 39. R06559 Polyhedrin of Spodoptera litu 249 4 5 2.40 0
 40. F94369 Fusion protein cong. 41kd pr 252 4 5 2.40 0
 41. R06201 O-acetylserine sulphyrylase (253 4 5 2.40 0
 42. R13274 Petunia extracellular chitina 254 4 5 2.40 0
 43. R22959 Human proteasome component HC 255 4 5 2.40 0
 44. R21516 Polyfunctional protease C8. 255 4 5 2.40 0
 45. R10169 Vaccinia virus 35k gene produc 258 4 5 2.40 0

1. US-08-121-713B-17 (1-6)
 R05171 Natural killer cell stimulatory factor (NKSF).

ID R05171 standard; protein; 291 AA.
 AC R05171;
 DT 11-OCT-1990 (first entry)
 DE Natural killer cell stimulatory factor (NKSF).
 KW Natural killer cell stimulatory factor; NKSF; cancer;
 KW leukaemia; anaemia; radiation; bone marrow transplantation; ds.
 OS Homo sapiens.
 PN WO9005147-A.
 PD 17-MAY-1990.
 PF 9-NOV-1989; 005027.
 PR 10-NOV-1988; US-269945.
 PA (GENE-) Genetics Inst Inc, (WIST-) Wistar Inst.
 PI Trinchieri G, Perussia B, Kobayashi M, Clark SC, Wong GG;
 DR WPI; 90-178800/23.
 DR N-PSDB; Q04701.
 PT New natural killer cell stimulatory factor -

PT useful for treating cancer and eg. leukopenia, bacterial and
 PT viral infections, anaemia and B- or T-cell deficiencies.
 PS Claim 17; Page 61; 69pp; English.
 CC Natural killer cell stimulatory factor (NKSF) is useful in
 CC treating cancer and other disease states which respond to NK cell
 CC activity. Such states may arise from immune or haematopoietic-cell
 CC deficiency following a bone marrow transplantation, exposure to
 CC radiation and disease.
 SQ Sequence 291 AA;
 SQ 12 A; 9 R; 7 N; 19 D; 0 B; 9 C; 12 Q; 23 E; 0 Z; 16 G; 6 H;
 SQ 12 I; 26 L; 22 K; 3 M; 10 F; 14 P; 29 S; 20 T; 9 W; 10 Y; 23 V;
 Initial Score = 5 Optimized Score = 5 Significance = 3.20
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X X
 FXREXA
 I I I
 KDKQEPKNKTFLRCEAKNYSGRFTCW
 120 130 X 140

2. US-08-121-713B-17 (1-6)
 R12493 Cytotoxic lymphocyte maturation factor (40 kD subu

ID R12493 standard; Protein; 313 AA.
 AC R12493;
 DT 10-SEP-1991 (first entry)
 DE Cytotoxic lymphocyte maturation factor (40 kD subunit).
 KW Cytotoxic lymphocyte maturation factor; CLMF; probe; antitumour;
 KW monoclonal antibodies; T cell.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= sig_peptide
 FT Protein 23..313
 FT /label= mat_protein
 FN EP-433827-A.
 FD 26-JUN-1991.
 PF 09-DEC-1990; 123670.
 PR 22-DEC-1989; US-455708.
 PR 09-MAY-1990; US-520935.
 PR 27-AUG-1990; US-572284.
 PA (HOFF) HOFFMANN-LA ROCHE AG.
 PI Chizsonite RA, Gately MK, Gubler UA, Hulmes JD, Pan YCE;
 PI Podlaski FJ, Stern AS;
 DR WPI; 91-186715/26.
 DR N-PSDB; Q12149.
 PT New cytotoxic lymphocyte maturation factor - used as antitumour
 PT agent and monoclonal antibodies used in transplantation to block
 PT proliferation and activation of cytotoxic T cells
 PS Disclosure; Fig 25 (a-c); 90pp; English.
 CC The natural CLMF protein is a 75 kD heterodimer comprising two
 CC polypeptide subunits, a 40 kD and a 35 kD subunit (Q12150).
 CC The subunits are bonded together via one or more disulfide bonds.
 CC The CLMF can interact with IL-2 to synergistically induce the
 CC cytolytic activity of Lymphokine Activated Killer (LAK) cells. It is
 CC also capable of stimulating T-cell growth. It can be used for the
 CC prepn. of LAK- and T-cell activator and antitumour compns. and for

CC stimulating Natural Killer (NK) cells.
SQ Sequence 313 AA;
SQ 15 A; 12 R; 7 N; 21 D; 0 B; 10 C; 13 Q; 23 E; 0 Z; 15 G; 6 H;
SQ 13 I; 26 L; 26 K; 3 M; 11 F; 14 P; 31 S; 23 T; 9 W; 10 Y; 25 V;
Initial Score = 5 Optimized Score = 5 Significance = 3.20
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X X
FXREXA
| | |
KDKQEPKNTFLRCEAKNYSGRFTCW
120 130 X 140

3. US-08-121-713B-17 (1-6)
R22769 Natural Killer cell stimulatory factor.

ID R22769 standard; Protein; 328 AA.
AC R22769;
DT 16-SEP-1992 (first entry)
DE Natural killer cell stimulatory factor.
KW NKSF; cytokine; gamma interferon; IFN; peripheral blood lymphocytes;
KW IL-2; GM-CSF; granulocyte macrophage colony stimulating factor; PBL;
ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 23..27
FT /label= tryptic peptide 3
FT Peptide 75..79
FT /label= tryptic fragment 1
FT Peptide 219..224
FT /label= tryptic fragment 2
FT Peptide 303..308
FT /label= tryptic fragment 7
FT Peptide 127..130
FT /label= tryptic fragment 8
FT Peptide 231..239
FT /label= tryptic fragment 9
FT Protein 23..328
FT /label= mature 40kD NKSF
FT Peptide 1..22
FT /label= signal peptide
FT W09205256-A.
PD 02-APR-1992.
PF 04-SEP-1991; U06332.
PR 18-SEP-1990; US-584941.
PA (GENE-) GENETICS INST INC.
PA (WIST-) WISTAR INST.
PI Clark S C, Hewick R, Kobayashi M, Perussia B, Trinchieri G;
PI Wong G G;
PI WPI; 92-132124/16.
DR N-PSDB; Q23586.
PT New natural killer cell stimulatory factor - useful as an
PT immunostimulant for inducing gamma-interferon and GM-CSF prodn.
PT in human peripheral blood lymphocytes
PS Claim 12; Page 23; 79pp; English.
CC NKSF is capable of inducing the production of gamma interferon in
CC human peripheral blood lymphocytes. It has subunits of 40 (R22769)

CC and 30-35 kD (R23729). It is claimed that NKSF is useful in the
CC treatment of bacterial and viral infections e.g. AIDS. It has a
CC specific activity of >1*10power7 dilution units in the gamma-IFN
CC induction assay. Additionally it has biological activity in GM-CSF
CC inducing assay; biological activity in activating natural killer
CC cells to kill leukaemia and tumour derived cells; biological activity
CC in a TNF induction assay using PHA activated T-lymphocytes; and co-
CC mitogenic activity on PBL. It also synergises with IL-2 in inducing
CC gamma IFN production in PBLs and maintaining PBL proliferation.
SQ Sequence 328 AA;
SQ 16 A; 12 R; 7 N; 21 D; 0 B; 11 C; 13 Q; 24 E; 0 Z; 15 G; 6 H;
SQ 13 I; 26 L; 26 K; 3 M; 11 F; 15 P; 37 S; 23 T; 11 W; 12 Y; 26 V;
Initial Score = 5 Optimized Score = 5 Significance = 3.20
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X X
FXREXA
| | |
KDKQEPKNTFLRCEAKNYSGRFTCW
120 130 X 140

4. US-08-121-713B-17 (1-6)
R45936 A glycosyltransferase.

ID R45936 standard; Protein; 365 AA.
AC R45936;
DT 26-JUL-1994 (first entry)
DE A glycosyltransferase.
KW Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
KW surface; oligosaccharide.
OS Homo sapiens.
FN W09402616-A.
PD 03-FEB-1994.
PF 20-JUL-1993; U06703.
PR 20-JUL-1992; US-914281.
PA (UNMI) UNIV MICHIGAN.
PI Lowe JB;
PI WPI; 94-048874/06.
DR N-PSDB; Q56908.
PT DNA fragment encoding a glycosyltransferase - can be used for in
PT vitro reactions to modify cell surface oligosaccharide(s) e.g.
PT blood gp. determinants, to protect against transplant rejection
PS Disclosure; Fig 3; 249pp; English.
CC The sequence is that of a human glycosyl transferase. The enzyme
CC may be non glycosylated. This prevents premature loss of enzyme
CC activity. It can also be used in in vitro reactions to modify cell
CC surface oligosaccharide mols. e.g. blood group determinants.
CC See also R45933-9.
SQ Sequence 365 AA;
SQ 32 A; 25 R; 9 N; 18 D; 0 B; 8 C; 16 Q; 15 E; 0 Z; 26 G; 14 H;
SQ 12 I; 44 L; 7 K; 9 M; 24 F; 24 P; 19 S; 22 T; 12 W; 7 Y; 22 V;
Initial Score = 5 Optimized Score = 5 Significance = 3.20
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

FXREXA
|||
QIRREFTLHDHLEEAQSVLGRLG
180 190 X 200

5. US-08-121-713B-17 (1-6)
R13751 GDP-Fuc:beta-D-galactoside alpha(1,2)-fucosyltrans

AC R13751 standard; Protein; 365 AA.

DT 07-NOV-1991 (first entry)

DE GDP-Fuc:beta-D-galactoside alpha(1,2)-fucosyltransferase.

KW Glycosyltransferase.

OS Homo sapiens.

PN WO9112340-A.

PD 22-AUG-1991.

PF 14-FEB-1991; U00899.

PR 14-FEB-1990; US-480133.

PR 14-FEB-1990; US-479858.

PR 12-DEC-1990; US-627621.

PA (UNMI) UNIV OF MICHIGAN.

PI Lowe JB.

DR MPI; 91-267151/36.

DR P-PSDB; R13751.

PT Isolation of gene conveying post-translational characteristic -

PT e.g. the presence of soluble or membrane bound oligo or

PT polysaccharide or glycosyltransferase.

PS Disclosure; Fig 3; 15pp; English.

CC The amino acid sequence codes for a protein capable of functioning

CC as a GDP- Fuc:beta-D-Gal alpha(1,2)- fucosyltransferase. The

CC enzymatic protein is represented by amino acids 33 to 365. The

CC function produced by the DNA sequence can be used in enzymatic

CC fucosylation of chain-terminating galactose residues on lactose-

CC amine or neolacto type beta-D-galactoside to alpha-2-L-fucose

CC residues. See also R13749-R13752.

SQ Sequence 365 AA;

SQ 31 A; 25 R; 9 N; 18 D; 0 B; 8 C; 16 Q; 15 E; 0 Z; 26 G; 14 H;

SQ 12 I; 44 L; 7 K; 9 M; 24 F; 25 P; 19 S; 22 T; 12 W; 7 Y; 22 V;

Initial Score = 5 Optimized Score = 5 Significance = 3.20

Residue Identity = 50% Matches = 3 Mismatches = 3

Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
|||

QIRREFTLHDHLEEAQSVLGRLG
180 190 X 200

6. US-08-121-713B-17 (1-6)
R33744 XR2.

ID R33744 standard; Protein; 440 AA.

AC R33744;

DT 23-JUL-1993 (first entry)

DE XR2.

KW Receptor; XR2; DNA binding domain; human; retinoic acid receptor-alpha;

KW hRAR-alpha; thyroid receptor-beta; hTR-beta; glucocorticoid receptor;

KW hGR; retinoid X receptor-alpha; hRXR-alpha; verht3; verhr5; ligand;
KW transcription-activation; response element.

OS Homo sapiens.

PN WO9306215-A.

PD 01-APR-1993.

PF 08-SEP-1992; U07570.

PR 17-SEP-1991; US-761068.

PA (SALK) SALK INST BIOLOGICAL STUDIES.

PI Borgmeyer UK, Evans RM, Giguere V, Mangelsdorf DJ;

PI Ong ES, Oro AE, Yao TP;

DR MPI; 93-117536/14.

DR N-PSDB; Q39088.

PT DNA encoding proteins of thyroid-steroid hormone receptor

PT super-family - useful for screening for agonists-antagonists of

PT such receptors

PS Claim 9; Page 40-42; 71pp; English.

CC This sequence represents the receptor XR2. This polypeptide has a

CC DNA binding domain comprising approx. 55% amino acid sequence identity

CC with the DNA binding domain of human retinoic acid receptor-alpha

CC (hRAR-alpha), about 56% amino acid sequence identity with the DNA

CC binding domain of human thyroid receptor-beta (hTR-beta), about 50%

CC amino acid sequence identity with the DNA binding domain of human

CC glucocorticoid receptor (hGR) and about 52% amino acid sequence

CC identity with the DNA binding domain of human retinoid X receptor-

CC alpha (hRXR-alpha). XR2 receptor can be used for testing a compound

CC for its ability to regulate transcription-activating effects of a

CC receptor polypeptide, identifying compounds which act as ligands for

CC the receptor polypeptides and for identifying response elements for

CC the receptor polypeptides.

SQ Sequence 440 AA;

SQ 36 A; 33 R; 9 N; 18 D; 0 B; 12 C; 29 Q; 35 E; 0 Z; 19 G; 15 H;

SQ 20 I; 46 L; 21 K; 13 M; 18 F; 33 P; 37 S; 14 T; 4 W; 7 Y; 21 V;

Initial Score = 5 Optimized Score = 5 Significance = 3.20

Residue Identity = 50% Matches = 3 Mismatches = 3

Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
|||

SSQAQGGSSCILIREATWPHSAGGTA
30 40 X 50

7. US-08-121-713B-17 (1-6)

R48205 H.tuberosus cytochrome P450 Cinnamate 4-hydroxylas

ID R48205 standard; Protein; 505 AA.

AC R48205;

DT 12-JUL-1994 (first entry)

DE H.tuberosus cytochrome P450 Cinnamate 4-hydroxylase.

KW Cinnamate 4-hydroxylase; Jerusalem artichoke; common white variety;

KW CADH; EC.1.14.13.1; cytochrome P450; mono oxidation; hydroxylation;

KW NADPH-cytochrome P450 reductase.

OS Helianthus tuberosus.

PN WO9401564-A.

PD 20-JAN-1994.

PF 02-JUL-1993; F00676.

PR 03-JUL-1992; FR-008254.

PA (ORSA-) ORSAN.

PI Kazmaier M, Mignotte VC, Pompon D, Renaud M, Teutsch H;
 PI Werck-Reichart D, Mignotte Vieux C, Werck-Reichart D;
 DR WPI; 94-035069/04.
 DR N-PSDB; Q55607.
 PT New yeast strain expressing plant cytochrome P450 mono:oxygenase
 PT - and simultaneously NADPH cytochrome reductase, with gene
 PT integrated, useful for bio conversion, esp. specific
 PT hydroxylation of cinnamate to coumarate
 PS Example 1; Fig 1; 71pp; French.
 CC The cytochrome P450 cinnamate 4-hydroxylase R48205 catalyses the
 CC hydroxylation of trans-cinnamate in the Jerusalem artichoke.
 CC The C4MH gene can be used in the production of yeast strains able
 CC to co-express mono-oxygenase (i.e. hydroxylation) activity of a
 CC plant-derived cytochrome P450 as well as NADPH cytochrome 450
 CC reductase activity.
 SQ Sequence 505 AA;
 SQ 31 A; 33 R; 27 N; 26 D; 0 B; 3 C; 18 Q; 37 E; 0 Z; 32 G; 10 H;
 SQ 38 I; 58 L; 38 K; 12 M; 27 F; 28 P; 15 S; 18 T; 8 W; 11 Y; 35 V;
 SQ 38 I; 58 L; 38 K; 12 M; 27 F; 28 P; 15 S; 18 T; 8 W; 11 Y; 35 V;
 Initial Score = 5 Optimized Score = 5 Significance = 3.20
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXEXA
 I I
 KPPEFRPFRFLEEEAKVANGNDFR
 420 X 430

8. US-08-121-713B-17 (1-6)
 R44140 Murine FACC encoded by cDNA clone pmfac2.
 ID R44140 standard; Protein; 558 AA.
 AC R44140;
 DT 20-MAY-1994 (first entry)
 DE Murine FACC encoded by cDNA clone pmfac2.
 KW Fanconi Anemia Group C; FACC; complementing cDNA; variant; diagnosis;
 KW open reading frame; Fanconi anemia; gene therapy.
 OS Mus musculus.
 PN W09322435-A.
 PD 11-NOV-1993.
 PF 27-APR-1993; CA0178.
 PR 29-APR-1992; US-876285.
 PR 21-JUL-1992; US-918313.
 PR 15-JAN-1993; US-003963.
 PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
 PA (UNME-) UNITED MEDICAL & DENTAL SCHOOL GUYS.
 PI Buchwald M, Mathew CG, Strathdee CA, Wevrick R;
 DR WPI; 93-368794/46.
 DR N-PSDB; Q51455.
 PT Human cDNA which complements Fanconi Anaemia gp. C - used to
 PT develop prods. for use in diagnosis, study and therapy of Fanconi
 PT Anaemia
 PS Claim 1; Fig 11; 137pp; English.
 CC The sequences given in R44140 and R51270 are encoded by murine cDNA
 CC variants from the Fanconi Anemia Group C Complementing (FACC) cDNA.
 CC These two cDNA molecules are cellular variants of a single cDNA
 CC transcribed from the same gene. The second cDNA contains an additional
 CC exon which was isolated in the clone pmfac7. The FACC proteins encoded
 CC by these two clones may be used for the diagnosis and study of Fanconi
 CC anemia and the cDNAs may be used in gene therapy.
 SQ Sequence 591 AA;
 SQ 46 A; 33 R; 9 N; 18 D; 0 B; 21 C; 30 Q; 52 E; 0 Z; 23 G; 19 H;
 SQ 23 I; 90 L; 21 K; 15 M; 26 F; 36 P; 48 S; 24 T; 12 W; 11 Y; 34 V;
 Initial Score = 5 Optimized Score = 5 Significance = 3.20
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

CC by these two clones may be used for the diagnosis and study of Fanconi
 CC anemia and the cDNAs may be used in gene therapy.
 SQ Sequence 558 AA;
 SQ 45 A; 29 R; 9 N; 18 D; 0 B; 17 C; 29 Q; 50 E; 0 Z; 20 G; 17 H;
 SQ 21 I; 89 L; 20 K; 15 M; 25 F; 31 P; 47 S; 20 T; 11 W; 11 Y; 34 V;
 Initial Score = 5 Optimized Score = 5 Significance = 3.20
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXEXA
 I I
 GWVDLAVAEILLREAEFPAGLLWLL
 410 X 420

9. US-08-121-713B-17 (1-6)
 R51270 Murine FACC encoded by cDNA clone pmfac7.
 ID R51270 standard; Protein; 591 AA.
 AC R51270;
 DT 20-MAY-1994 (first entry)
 DE Murine FACC encoded by cDNA clone pmfac7.
 KW Fanconi Anemia Group C; FACC; complementing cDNA; variant; diagnosis;
 KW open reading frame; Fanconi anemia; gene therapy.
 OS Mus musculus.
 FH Key Location/Qualifiers
 FT Peptide 514..546
 FT /note="Encoded by the additional exon"
 PN W09322435-A.
 PD 11-NOV-1993.
 PF 27-APR-1993; CA0178.
 PR 29-APR-1992; US-876285.
 PR 21-JUL-1992; US-918313.
 PR 15-JAN-1993; US-003963.
 PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
 PA (UNME-) UNITED MEDICAL & DENTAL SCHOOL GUYS.
 PI Buchwald M, Mathew CG, Strathdee CA, Wevrick R;
 DR WPI; 93-368794/46.
 DR N-PSDB; Q51456.
 PT Human cDNA which complements Fanconi Anaemia gp. C - used to
 PT develop prods. for use in diagnosis, study and therapy of Fanconi
 PT Anaemia
 PS Claim 1; Fig 12; 137pp; English.
 CC The sequences given in R44140 and R51270 are encoded by murine cDNA
 CC variants from the Fanconi Anemia Group C Complementing (FACC) cDNA.
 CC These two cDNA molecules are cellular variants of a single cDNA
 CC transcribed from the same gene. The second cDNA contains an additional
 CC exon which was isolated in the clone pmfac7. The FACC proteins encoded
 CC by these two clones may be used for the diagnosis and study of Fanconi
 CC anemia and the cDNAs may be used in gene therapy.
 SQ Sequence 591 AA;
 SQ 46 A; 33 R; 9 N; 18 D; 0 B; 21 C; 30 Q; 52 E; 0 Z; 23 G; 19 H;
 SQ 23 I; 90 L; 21 K; 15 M; 26 F; 36 P; 48 S; 24 T; 12 W; 11 Y; 34 V;
 Initial Score = 5 Optimized Score = 5 Significance = 3.20
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
||
GWDLAVAEILLREAEPPAGLLWL
410 X 420

10. US-08-121-713B-17 (1-6)
R30497 N-terminal of LH receptor/FSH receptor chimera #6
- ID R30497 standard; protein; 757 AA.
AC R30497;
DE 10-MAY-1993 (first entry)
PT N-terminal of LH receptor/FSH receptor chimera #6.
KW Follicle stimulating hormone receptor; luteinising hormone receptor;
KW human chorionic gonadotrophin; glycoprotein hormone receptor;
KW chimera; chimera.
OS Chimaeric; homo sapiens.
PN W09222667-A.
PD 23-DEC-1992.
PF 12-JUN-1992; U04987.
PR 14-JUN-1991; US-715911.
PA (OYNE-) UNIV NEW JERSEY.
PI Bernard M, Moyle WR, Myers R;
DR WPI; 93-018150/02.
PT Glyco:protein hormone receptor analogues - having binding
PT affinity to human chorionic gonadotrophin, luteinising and
PT follicle stimulating hormones, useful in bio:immunoassays
PS Examples; Fig 12; 103pp; English.
CC This sequence represents the N-terminal of a novel protein having a
CC binding affinity for human chorionic gonadotrophin (hCG), luteinising
CC hormone (LH), and follicle stimulating hormone (FSH). The protein
CC itself is a chimera having residues from both the FSH receptor,
CC and LH receptor. The receptor analogues can be used in bioimmunoassays
CC for the simultaneous detection of both LH (or hCG) and FSH as
CC well as their ratio of biological activities. The analogues can also
CC be used for raising, purifying and assaying antibodies to the
CC analogues. Coding sequence for the chimera was produced by two step
CC PCR.
SQ Sequence 757 AA;
SQ 47 A; 31 R; 40 N; 26 D; 0 B; 27 C; 30 Q; 43 E; 0 Z; 43 G; 13 H;
SQ 50 I; 88 L; 36 K; 15 M; 42 F; 35 P; 70 S; 41 T; 9 W; 25 Y; 46 V;
Initial Score = 5 Optimized Score = 5 Significance = 3.20
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0
- X X
FXREXA
||
NPGEDIMGVAFLEDEAWVVGMAILMS
360 X 370 380
11. US-08-121-713B-17 (1-6)
R05935 Secreted GPIIb subunit of multiple subunit polypep
- ID R05935 standard; protein; 993 AA.
AC R05935;
DT 22-NOV-1990 (first entry)

- DE Secreted GPIIb subunit of multiple subunit polypeptide (MSP)
DE GPIIb-IIIa.
KW GPIIb; GPIIIa; MSP GPIIb-IIIa; anti-coagulant; anti-inflammatory;
KW immunosuppressant; ds.
OS Homo sapiens.
PN W09006953-A.
PD 28-JUN-1990.
PF 20-DEC-1989; U05743.
PR 22-DEC-1988; US-290224.
PR 01-DEC-1989; US-444490.
PA (GETH) GENENTECH INC.
PI BODARY SC, GORMAN CM, NAPIER MA, MCLEAN JW;
DR WPI; 90-224496/29.
DR N-PSDB; Q05271.
PT Soluble analogues of multi-sub-unit polypeptide prodn. - by
PT transforming host cells with nucleic acid modified to prevent
PT membrane attachment, useful in diagnosis and ligand purific.
FS Disclosure; 47pp; English.
CC Peptide product is analogous to MSP with better stability, it is
CC not membrane bound and so may be collected as a cell secretion from
CC a transformed host. The product is useful as a diagnostic reagent
CC eg. immunoassay of MSP; in purification of ligands and matrix
CC proteins; and therapeutically as agonists or antagonists of the
CC corresponding membrane bound receptor.
SQ Sequence 993 AA;
SQ 81 A; 52 R; 36 N; 49 D; 0 B; 20 C; 62 Q; 54 E; 0 Z; 87 G; 23 H;
SQ 26 I; 122 L; 22 K; 12 M; 33 F; 73 P; 75 S; 36 T; 17 W; 28 Y; 85 V;
Initial Score = 5 Optimized Score = 5 Significance = 3.20
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0
- X X
FXREXA
||
HSPICHTTMFLRDEADFROKLSPIV
580 X 590
12. US-08-121-713B-17 (1-6)
R44430 eryA region polypeptide module #1.
- ID R44430 standard; Protein; 3398 AA.
AC R44430;
DT 22-DEC-1993 (first entry)
DE eryA region polypeptide module #1.
KW Saccharopolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW erythromycin; condensation; elongation; acyl chain growth;
KW gene replacement.
OS Saccharopolyspora erythraea.
PN W09313663-A.
PD 22-JUL-1993.
PF 17-JAN-1992; U00427.
PR 17-JAN-1992; WO-U00427.
PA (ABBO) ABBOTT LAB.
PI Donadio S, Katz L, McAlpine JB;
DR WPI; 93-242804/30.
DR N-PSDB; Q46806.
PT Biosynthesis of specific polyketide analogues esp. erythromycin
PT cpds. - by introducing altered biosynthetic gene-contg. DNA into

PT microorganisms
 PS Disclosure; Fig 2; 133pp; English.
 CC The sequences given in R4430-32 are encoded by the *eryA* fragment of
 CC the *Saccaropolyspora erythraea* genome. These polypeptides are
 CC involved in the biosynthesis of the polyketide segment of erythromycin.
 CC *eryA* is organised in modules and each module takes care of one
 CC condensation step. The precise succession of elongation steps is
 CC dictated by the genetic order of the modules. The DNA encoding
 CC these polypeptides may be specifically altered such that novel
 CC polyketide molecules of desired structure are produced. Three types
 CC of alteration may be produced; those inactivating a single function in
 CC a module which does not arrest acyl chain growth; those inactivating a
 CC single function in a module which does affect chain growth; and those
 CC affecting an entire module. The mutations may be introduced by gene
 CC replacement.

SQ Sequence 3398 AA;
 SQ 554A; 288R; 39 N; 201D; 0 B; 29 C; 81 Q; 239E; 0 Z; 349G; 66 H;
 SQ 72 I; 331L; 22 K; 44 M; 89 F; 193P; 208S; 171T; 56 W; 34 Y; 332V;

Initial Score = 5 Optimized Score = 5 Significance = 3.20
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXREXA
 |||
 EPHLDFEVPFLRAEAAREQDAALS
 980 X 990

13. US-08-121-713B-17 (1-6)
 R38556 Sequence of peptide which corresponds to ragweed p

ID R38556 standard; peptide; 15 AA.
 AC R38556;
 DT 08-NOV-1993 (first entry)
 DE Sequence of peptide which corresponds to ragweed pollen antigen3
 DE Ra3.
 KW Allergen epitope; ragweed pollen; allergy; graft rejection.

OS Synthetic.
 PN W09312145-A.
 PD 24-JUN-1993.
 PF 18-DEC-1992; U11238.
 PR 19-DEC-1991; US-811050.
 PA (BAYD) BAYLOR COLLEGE MEDICINE.
 PI Ashizawa T, Acaasi MZ;
 DR WPI; 93-214099/26.

PT Immune disorders treatment reagent prodrn. for ragweed pollen
 PT allergy - by forming peptide contg. allergen epitope with polymer
 PT e.g. polyethylene glycol, deriving aminoacid, de-protecting,
 PT cleaving, and purifying for Graves disease
 PS Claim 4; Figure 1 no.2; 76pp; English.
 CC The peptides of the invention are allergen epitopes bonded to a
 CC tolerogenic polymer (TP) such as monomethoxypolyethylene glycol (mPEG)
 CC or polyvinyl alcohol (PVA). Pref.the peptide is attached to the
 CC resin at the C-terminus and TP attached to the N-terminus. It is
 CC used for treating a patient having, or likely to develop, ragweed
 CC allergy.
 SQ Sequence 15 AA;
 SQ 2 A; 1 R; 0 N; 2 D; 0 B; 1 C; 0 Q; 3 E; 0 Z; 0 G; 1 H;

SQ 1 I; 0 L; 1 K; 0 M; 0 F; 0 P; 0 S; 0 T; 1 W; 1 Y; 1 V;
 Initial Score = 4 Optimized Score = 5 Significance = 2.40
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXREXA
 |||
 EVMREEAVHACDIKD
 X X 10

14. US-08-121-713B-17 (1-6)
 R41497 TNF inhibitory peptide IX.

ID R41497 standard; peptide; 21 AA.
 AC R41497;
 DT 23-FEB-1994 (first entry)
 DE TNF inhibitory peptide IX.
 KW Tumour necrosis factor; TNF; inhibition; solid phase synthesis; ss.
 OS Synthetic.
 PN J05194594-A.
 PD 03-AUG-1993.
 PF 21-JAN-1992; JP-029044.
 PR 21-JAN-1992; JP-029044.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 DR WPI; 93-282916/36.
 PT TNF inhibitory novel peptide(s) - include N-terminal amino Gp.
 PT which is opt. modified with acetyl, t-butoxy-carbonyl or
 PT benzyl-oxy-carbonyl Gp. and C-terminal carboxy Gp. is opt.
 PT amidated
 PS Claim 1; Page 6; 8pp; Japanese.
 CC The sequences given in R41489-99 are tumour necrosis factor (TNF)
 CC inhibitory peptides. They may optionally be modified at the N-
 CC terminal with an acetyl, t-butoxycarbonyl or benzylloxycarbonyl, and
 CC at the C-terminal they are optionally amidated. These peptides are
 CC produced by solid phase synthesis methods and may be produced at low
 CC cost.

SQ Sequence 21 AA;
 SQ 3 A; 1 R; 2 N; 0 D; 0 B; 0 C; 2 Q; 3 E; 0 Z; 2 G; 1 H;
 SQ 0 I; 1 L; 1 K; 0 M; 2 F; 0 P; 0 S; 2 T; 0 W; 0 Y; 1 V;

Initial Score = 4 Optimized Score = 5 Significance = 2.40
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXREXA
 |||
 NTVATAHAGFFELRENEG
 10 X 20

15. US-08-121-713B-17 (1-6)
 R25980 Peptide monomer 18.

ID R25980 standard; Protein; 23 AA.
 AC R25980;
 DT 21-JAN-1993 (first entry)

Results file sql7pir.res made by maryh on Fri 19 May 95 10:47:00-PDT.

SCORE 0| 1 |1| 2 | 2 | 3 | 3 | 4 | 4 | 5
STDEV -1 0 0

PARAMETERS

Similarity matrix Unitary K-tuple 2
Mismatch penalty 1 Joining penalty 20
Gap penalty 1.00 Window size 6
Gap size penalty 0.05
Cutoff score 0
Randomization group 0

SEARCH STATISTICS

Scores: Mean 2 Median 3 Standard Deviation 1.08

Times: CPU Total Elapsed
00:01:10.99 00:01:11.00

Number of residues: 22468834
Number of sequences searched: 75511
Number of scores above cutoff: 4422

The scores below are sorted by initial score.
Significance is calculated based on initial score.
A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Init. Opt.
*** 2 standard deviations above mean ***				
1. C49958	AMP-PK=AMP-activated protein	58	5	5
2. A41960	homoserine O-succinyltransfer	69	5	5
3. S31563	hypothetical protein YGL024 -	90	5	5
4. HRIN	hemerythrin - sipunculid (Sip	113	5	5
5. S13197	profilin - sea urchin (Clypea	139	5	5
6. S13198	profilin - sea urchin (Anthoc	139	5	5
7. A44777	profilin spCoell - sea urchin	142	5	5
8. S17791	aminoacyl-tRNA hydrolase (EC	194	5	5
9. S16753	aminoacyl-tRNA hydrolase (EC	194	5	5
10. S05446	tropomyosin, nonmuscle - Japa	221	5	5
11. B32014	traJ protein - Escherichia co	223	5	5
12. S44924	39K antigen - Entamoeba histo	237	5	5
13. S44923	39K antigen - Entamoeba histo	237	5	5
14. S24403	tropomyosin alpha - chicken	245	5	5
15. A60364	tropomyosin - migratory locus	283	5	5
16. C25242	tropomyosin, exon 9D - fruit	285	5	5
17. A25561	tropomyosin II, muscle - frui	285	5	5
18. B25242	tropomyosin, exon 9C - fruit	286	5	5
19. A25242	tropomyosin, exon 9B - fruit	286	5	5
20. DMHY	desmin - golden hamster (frag	298	5	5
21. S21688	interleukin-12 - human	306	5	5
22. XYECM	homoserine O-succinyltransfer	309	5	5

23. BWSMNG	strN protein - Streptomyces g	319	5	5	2.79	0
24. B39359	cytotoxic lymphocyte maturati	328	5	5	2.79	0
25. S46819	hypothetical protein YHL009c	330	5	5	2.79	0
26. S17823	protechlorophyllide reductase	333	5	5	2.79	0
27. B49850	chlorin reductase subunit Bch	333	5	5	2.79	0
28. S44922	39K antigen - Entamoeba histo	335	5	5	2.79	0
29. S35250	hrpN protein - Pseudomonas so	357	5	5	2.79	0
30. A36047	galactoside 2-alpha-L-fucosyl	365	5	5	2.79	0
31. DCECIS	isocitrate dehydrogenase (NAD	416	5	5	2.79	0
32. S36799	calreticulin (clone 9,4) - bo	421	5	5	2.79	0
33. A49424	patterning protein Sonic hedg	425	5	5	2.79	0
34. S15471	EAV element protein - Chicken	433	5	5	2.79	0
35. B53193	hedgehog homolog vhh-1 - rat	437	5	5	2.79	0
36. A54104	desmin - mouse	469	5	5	2.79	0
37. A24783	desmin - hamster	469	5	5	2.79	0
38. A33312	sarcoplasmic reticulum 53K gl	472	5	5	2.79	0
39. S28495	trans-cinnamate 4-monooxygena	505	5	5	2.79	0
40. A17454	trans-cinnamate 4-monooxygena	505	5	5	2.79	0
41. S19266	anthranilate synthase (EC 4.1	508	5	5	2.79	0
42. DESMP	IMP dehydrogenase (EC 1.1.1.2	513	5	5	2.79	0
43. S20589	Transcription factor TFIIF la	517	5	5	2.79	0
44. S20248	Transcription factor TFIIF la	517	5	5	2.79	0
45. A38101	potassium channel HLK-3, volt	523	5	5	2.79	0

1. US-08-121-713B-17 (1-6)

AMP-PK=AMP-activated protein kinase 63 kda catalyt

ENTRY C49958 #type fragment
TITLE AMP-PK=AMP-activated protein kinase 63 kda catalytic subunit
- pig (fragment)
ORGANISM #formal name Sus scrofa domestica #common name domestic pig
DATE 06-Oct-1994 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994

ACCESSIONS C49958
REFERENCE Mitchell, K.I.; Stapleton, D.; Gao, G.; House, C.;
#authors Michel, B.; Katsis, F.; Witters, L.A.; Kemp, B.E.
#journal J. Biol. Chem. (1994) 269:2361-2364
#title Mammalian AMP-activated protein kinase shares structural and
functional homology with the catalytic domain of yeast Snfl
protein kinase.
#accession C49958
#status preliminary
#molecule_type protein
#residues 1-58 #label MIT
#cross-references NCBI:143412
#note sequence extracted from NCBI backbone

SUMMARY #length 58 #checksum 6573

SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.79
Residue Identity = 66% Matches = 4 Mismatches = 2
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
| | |
GLSNMMSDGEFLRXAXXPVAAPEV
30 X 40

```
2. US-08-121-713B-17 (1-6)
A41960 homoserine O-succinyltransferase (EC 2.3.1.46) - S

ENTRY      A41960      #type fragment
TITLE      homoserine O-succinyltransferase (EC 2.3.1.46) - Salmonella
            typhimurium (fragment)
ORGANISM    #formal name Salmonella typhimurium
DATE        31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
            31-Dec-1993
ACCESSIONS  A41960
REFERENCE   A41960
#authors    Mares, R.; Urbanowski, M.L.; Stauffer, G.V.
#journal     J. Bacteriol. (1992) 174:390-397
#title       Regulation of the Salmonella typhimurium metA gene by the
            metR protein and homocysteine.
#cross-references MVID:92105004
#accession   A41960
##status     preliminary
##molecule type DNA
##residues   1-69 ##label MAR
##cross-references NCBI:75596; NCBI:75598
##note       sequence extracted from NCBI backbone
KEYWORDS    acyltransferase
SUMMARY     #length 69 #checksum 2217
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.79
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
| |
RVLDLPVNFLEENFVDMTTRAS
10 X 20

3. US-08-121-713B-17 (1-6)
S31563 hypothetical protein YGL024 - yeast (Saccharomyces

ENTRY      S31563      #type complete
TITLE      hypothetical protein YGL024 - yeast (Saccharomyces
            cerevisiae)
ORGANISM    #formal name Saccharomyces cerevisiae
DATE        28-May-1993 #sequence_revision 28-May-1993 #text_change
            18-Nov-1994
ACCESSIONS  S31563
REFERENCE   S15040
#authors     Chen, W.; Balzi, E.; Capieaux, E.; Choder, M.; Goffeau, A.
#journal     Yeast (1991) 7:287-299
#title       The DNA sequencing of the 17 kb HindIII fragment spanning the
            LEU1 and ATE1 loci on chromosome VII from Saccharomyces
            cerevisiae reveals the PDR6 gene, a new member of the
            genetic network controlling pleiotropic drug resistance.
#cross-references MVID:91353083
#accession   S31563
##molecule type DNA
##residues   1-90 ##label CHE
```

```
##cross-references GB:S58126
##note       translation of nucleotide sequence not given
GENETICS
#map_position 7L
SUMMARY      #length 90 #molecular-weight 10527 #checksum 7540
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.79
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
| |
MKNLLINYLREAEAFQLEMLFCSG
10 X 20

4. US-08-121-713B-17 (1-6)
HRIN hemerythrin - sipunculid (Siphonosoma cumanense)

ENTRY      HRIN      #type complete
TITLE      hemerythrin - sipunculid (Siphonosoma cumanense)
ORGANISM    #formal name Siphonosoma cumanense
DATE        30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
            30-Jun-1993
ACCESSIONS  JT0556
REFERENCE   JT0556
#authors     Ochida, T.; Yano, H.; Satake, K.; Kubota, I.; Tsugita, A.
#journal     Protein Seq. Data Anal. (1990) 3:141-147
#title       The amino acid sequence of hemerythrin from Siphonosoma
            cumanense.
#cross-references MVID:90301732
#accession   JT0556
##molecule type protein
##residues   1-113 ##label UCH
##note       3-Glu, 10-Asp, 60-Gly, 66-Asn, and 83-Gln were also
            found
COMMENT      Hemerythrin is a respiratory protein found in several phyla of
            marine invertebrates. The functional molecule is a trimer
            composed of homologous chains, each of which contains a pair of
            nonheme iron atoms that reversibly bind molecular oxygen.
CLASSIFICATION #superfamily hemerythrin
FEATURE
25,54,58,73,77,101, #binding_site 2Fe-O cluster (His, His, Glu, His, His,
106 His, Asp) #status predicted
SUMMARY      #length 113 #molecular-weight 12437 #checksum 1365
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.79
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
| |
DNVAVACLVAHFLFEAAAMQVAKYGGY
50 X 60 70
```

5. US-08-121-713B-17 (1-6)
 S13197 profilin - sea urchin (Clypeaster japonicus)

ENTRY S13197 #type complete
TITLE profilin - sea urchin (Clypeaster japonicus)
ORGANISM #formal_name Clypeaster japonicus
DATE 02-Dec-1993; #sequence_revision 02-Dec-1993; #text_change
 02-Dec-1993

ACCESSIONS S13197
REFERENCE #authors
 Takagi, T.; Mabuchi, I.; Hosoya, H.; Furuhashi, K.; Hatano, S.
#journal Eur. J. Biochem. (1990) 192:777-781
#title Primary structure of profilins from two species of Echinoidea
 and Physarum polycephalum.
#cross-references MUID:91006174
#accession S13197
#status preliminary
#residues 1-139 #label TAK
SUMMARY #length 139 #molecular-weight 14540 #checksum 8410
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.79
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

 X X
 FXREXA
 | | |
GVHAEGLYQFLREEDAKLVLAKKKG
80 X 90

6. US-08-121-713B-17 (1-6)
 S13198 profilin - sea urchin (Anthocidaris crassispina)

ENTRY S13198 #type complete
TITLE profilin - sea urchin (Anthocidaris crassispina)
ORGANISM #formal_name Anthocidaris crassispina
DATE 02-Dec-1993; #sequence_revision 02-Dec-1993; #text_change
 02-Dec-1993

ACCESSIONS S13198
REFERENCE #authors
 Takagi, T.; Mabuchi, I.; Hosoya, H.; Furuhashi, K.; Hatano, S.
#journal Eur. J. Biochem. (1990) 192:777-781
#title Primary structure of profilins from two species of Echinoidea
 and Physarum polycephalum.
#cross-references MUID:91006174
#accession S13198
#status preliminary
#residues 1-139 #label TAK
SUMMARY #length 139 #molecular-weight 14569 #checksum 895
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.79
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

 X X
 FXREXA
 | | |
GIVADGTKYQFLREEDGKLVLAKKKG
80 X 90

7. US-08-121-713B-17 (1-6)
 A44777 profilin spCoell - sea urchin (Strongylocentrotus purpuratus)

ENTRY A44777 #type complete
TITLE profilin spCoell - sea urchin (Strongylocentrotus purpuratus)
ORGANISM #formal_name Strongylocentrotus purpuratus #common_name
 purple urchin
DATE 03-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
 30-Sep-1993

ACCESSIONS A44777
REFERENCE #authors
 Smith, L.C.; Britten, R.J.; Davidson, E.H.
 Mol. Biol. Cell (1992) 3:403-414
#journal SpCoell: a sea urchin profilin gene expressed specifically in
 coelomocytes in response to injury.
#accession A44777
#status preliminary
#molecule type mRNA
#residues 1-142 #label SMI
#cross-references GB:S42185
SUMMARY #length 142 #molecular-weight 15282 #checksum 3203
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.79
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

 X X
 FXREXA
 | | |
GIYVNGTKYQFLREEDSKLVLAKKKG
80 X 90

8. US-08-121-713B-17 (1-6)
 S17791 aminoacyl-tRNA hydrolase (EC 3.1.1.29) - Escherich

ENTRY S17791 #type complete
TITLE aminoacyl-tRNA hydrolase (EC 3.1.1.29) - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
 22-Nov-1993

ACCESSIONS S17791
REFERENCE #authors
 Garcia-Villegas, M.R.; de la Vega, F.M.; Galindo, J.M.;
 Segura, M.; Buckingham, R.H.; Guarneros, G.
#journal EMBO J. (1991) 10:3549-3555
#title Peptidyl-tRNA hydrolase is involved in lambda inhibition of
 host protein synthesis.
#cross-references MUID:92007806
#accession S17791
#status preliminary
#residues 1-194 #label GAR

```
##cross-references EMBL:X61941
SUMMARY   #length 194 #molecular-weight 21082 #checksum 9661
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.79
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
|||
VDLLAEELRLAPLREAAKFFGYTSRVT
30 40 X 50

9. US-08-121-713B-17 (1-6)
aminoacyl-tRNA hydrolase (EC 3.1.1.29) - Escherich
SI6753 #type complete
aminoacyl-tRNA hydrolase (EC 3.1.1.29) - Escherichia coli
TITLE #formal name Escherichia coli
ORGANISM #sequence_revision 21-Nov-1993; #text_change
DATE 21-Nov-1993
ACCESSIONS SI6753
REFERENCE SI6753
#authors Garcia-Villegas, R.; de la Vega, X.Y.Z.; Galindo, J.M.;
Segura, E.; Buckingham, R.; Guarneros, G.
#submission submitted to the EMBL Data Library, August 1991
#description Peptidyl-tRNA hydrolase is involved in lambda inhibition of
host Protein Synthesis.
#accession SI6753
#status preliminary
#residues 1-194 #label GAR
##cross-references EMBL:X61941
SUMMARY   #length 194 #molecular-weight 21050 #checksum 9670
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.79
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
|||
VDLLAEELRLAPLREAAKFFGYTSRVT
30 40 X 50

10. US-08-121-713B-17 (1-6)
tropomyosin, nonmuscle - Japanese quail (fragment)
S05446
ENTRY S05446 #type fragment
TITLE tropomyosin, nonmuscle - Japanese quail (fragment)
ORGANISM #formal name Coturnix coturnix japonica #common_name Japanese
quail
DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
28-Oct-1994
ACCESSIONS S05446
REFERENCE S05444
#authors Lindquester, G.J.; Flach, J.E.; Fleener, D.E.; Hickman, K.H.;
```

```
Devlin, R.B.
#journal Nucleic Acids Res. (1989) 17:2099-2118
#title Avian tropomyosin gene expression.
#cross-references MUID:89183613
#accession S05446
#molecule_type DNA
#residues 1-221 #label LIN
#cross-references EMBL:X16231
#note the authors translated the codon TCC for residue 51 as
Ala and GAG for residue 187 as Asp
GENETICS
#introns 44/3; 89/2; 128/3; 152/2; 177/3; 198/3
KEYWORDS alternative splicing
SUMMARY #length 221 #checksum 4781
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.79
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
|||
RLQREVDOERALRREAESEVASLNRR
30 40 X 50
> O <
O I JO IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sql7spt.res made by on Fri 19 May 95 8:57:53-PDT.

Query sequence being compared:US-08-121-713B-17 (1-6)
Number of sequences searched: 43470
Number of scores above cutoff: 4182

Results of the initial comparison of US-08-121-713B-17 (1-6) with:
Data bank : Swiss-Prot 31, all entries
```

```
100000-
-
N -
US0000-
M -
B -
E -
R -
O -
F10000-
S -
E 5000-
Q *
U -
E -
N -
```


43. ODO1 AZOVI 2-OXOGLUTARATE DEHYDROGENASE 943 5 5 3.38 0
 44. NOF DROME 112 KD PROTEIN IN NOF-FB TRAN 984 5 5 3.38 0
 45. ITAB_HUMAN PLATELET MEMBRANE GLYCOPROTEIN 1039 5 5 3.38 0

1. US-08-121-713B-17 (1-6)
 META_SALTY HOMOSERINE O-SUCCINYLTRANSFERASE (EC 2.3.1.46) (HO
 ID META_SALTY STANDARD; PRT; 69 AA.
 AC P374I3;
 DT 01-OCT-1994 (REL. 30, CREATED)
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE HOMOSERINE O-SUCCINYLTRANSFERASE (EC 2.3.1.46) (HOMOSERINE O-
 TRANSUCCINYLASE) (FRAGMENT).
 GN META.
 OS SALMONELLA TYPHIMURIDM.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 [1]
 RN SEQUENCE FROM N.A.
 RP 92105004
 RM 92105004
 RA MARES R., URBANOWSKI M.L., STAUFFER G.V.;
 RL J. BACTERIOL. 174:390-397(1992).
 CC -1- CATALYTIC ACTIVITY: SUCCINYL-COA + L-HOMOSERINE = COA +
 O-SUCCINYL-L-HOMOSERINE.
 CC -1- PATHWAY: FIRST ENZYME AND ONE OF THE KEY ENZYMES OF METHIONINE
 CC BIOSYNTHESIS PATHWAY.
 DR EMBL; M74188; STMETA.
 DR PIR; A41960; A41960.
 DR STYGENE; SG10222; META.
 KW METHIONINE BIOSYNTHESIS; TRANSFERASE; ACYLTRANSFERASE.
 FT NON TER 69 69
 SQ SEQUENCE 69 AA; 7976 MW; 24970 CN;

Initial Score = 5 Optimized Score = 5 Significance = 3.38
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXREXA
 | |
 RVLDLPVNFLEENVFDMTTSRAS
 10 X 20

2. US-08-121-713B-17 (1-6)
 HEMT_SIPC HEMERYTHRIN.
 ID HEMT_SIPC STANDARD; PRT; 113 AA.
 AC P22786;
 DT 01-AUG-1991 (REL. 19, CREATED)
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
 DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
 DE HEMERYTHRIN.
 OS SIPHONOSOMA CUMANENSE.
 OC EUKARYOTA; METAZOA; SIPUNCULA; SIPUNCULIDAE.
 [1]
 RN SEQUENCE.
 RP 90301732

UCHIDA T., YANO H., SATAKE K., KUBOTA I., TSUGITA A.;
 RL PROTEIN SEQ. DATA ANAL. 3:141-147(1990).
 CC -1- FUNCTION: HEMERYTHRIN IS A RESPIRATORY PROTEIN IN BLOOD CELLS OF
 CERTAIN MARINE WORMS. THE OXYGEN-BINDING SITE IN EACH CHAIN
 CC CONTAINS TWO IRON ATOMS.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SIMILARITY: TO OTHER HEMERYTHRINS AND TO MYOHEMERYTHRINS.
 DR PIR; JTO556; HRIN.
 DR HSSP; P02246; 2HMQ.
 DR PROSITE; PS00550; HEMERYTHRINS.
 KW OXYGEN TRANSPORT; METAL-BINDING; IRON.
 FT METAL 25 25 IRON 1 (BY SIMILARITY).
 FT METAL 54 54 IRON 1 (BY SIMILARITY).
 FT METAL 58 58 IRON 1 AND 2 (BY SIMILARITY).
 FT METAL 73 73 IRON 2 (BY SIMILARITY).
 FT METAL 77 77 IRON 2 (BY SIMILARITY).
 FT METAL 101 101 IRON 2 (BY SIMILARITY).
 FT METAL 106 106 IRON 1 AND 2 (BY SIMILARITY).
 FT VARIANT 3 3 P -> E.
 FT VARIANT 10 10 W -> D.
 FT VARIANT 60 60 A -> G.
 FT VARIANT 66 66 K -> N.
 FT VARIANT 83 83 K -> Q.
 SQ SEQUENCE 113 AA; 12437 MW; 61673 CN;

Initial Score = 5 Optimized Score = 5 Significance = 3.38
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXREXA
 | |
 DNAYACLVAHLEFEEAAMQVAKYGGY
 50 X 60 70

3. US-08-121-713B-17 (1-6)
 PROF_CLYJA PROFILIN.
 ID PROF_CLYJA STANDARD; PRT; 139 AA.
 AC P18321;
 DT 01-NOV-1990 (REL. 16, CREATED)
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
 DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
 DE PROFILIN.
 OS CLYPEASTER JAPONICUS (SAND DOLLAR).
 OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
 [1]
 RN SEQUENCE.
 RP 91006174
 RM 91006174
 RA TAKAGI T., MABUCHI I., HOSOVA H., FURUHASHI K., HATANO S.;
 RL EUR. J. BIOCHEM. 192:777-781(1990).
 CC -1- FUNCTION: PROFILIN PREVENTS THE POLYMERIZATION OF ACTIN.
 CC -1- OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC ACTIN
 CC IN A 1:1 RATIO.
 DR PIR; S13197; S13197.
 DR HSSP; P07763; IACF.
 DR PROSITE; PS00414; PROFILIN.
 KW ACTIN-BINDING; ACETYLATION.

FT MOD RES 1 1 ACETYLATION.
SQ SEQUENCE 139 AA; 14540 MW; 92443 CN;
Initial Score = 5 Optimized Score = 5 Significance = 3.38
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
| | |
GVHAEGLYQFLREEDAKVLAKKKG
80 X 90

4. US-08-121-713B-17 (1-6)
PROF_ANTCR PROFILIN.
ID PROF ANTCR STANDARD; PRT; 139 AA.
AC P18320;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
DE PROFILIN.
OS ANTHOCIDARIS CRASSISPINA (SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC PERISCHOECHINOIDEA.
RN [1]
RP SEQUENCE.
RM 91006174
RA TAKAGI T., MABUCHI I., HOSoya H., FURUHASHI K., HATANO S.;
RL EUR. J. BIOCHEM. 192:777-781(1990).
CC -!- FUNCTION: PROFILIN PREVENTS THE POLYMERIZATION OF ACTIN.
CC -!- OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC ACTIN
CC IN A 1:1 RATIO.
DR PIR; S13198; S13198.
DR HSSP; P07763; 2ACG.
DR PROSITE; PS00414; PROFILIN.
KW ACTIN-BINDING; ACETYLATION.
FT MOD RES 1 1 ACETYLATION.
SQ SEQUENCE 139 AA; 14569 MW; 95847 CN;

Initial Score = 5 Optimized Score = 5 Significance = 3.38
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
| | |
GIVADGTYQFLREEDGKVLAKKKG
80 X 90

5. US-08-121-713B-17 (1-6)
PRO1_STRPU PROFILIN.
ID PRO1 STRPU STANDARD; PRT; 141 AA.
AC P32006;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)

DE PROFILIN.
GN SPCOEL1.
OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC EUECHINOIDEA.
RN [1]
RP SEQUENCE FROM N.A.
RM 92360908
RA SMITH L.C., BRITTEN R.J., DAVIDSON E.H.;
RL MOL. BIOL. CELL 3:403-414(1992).
CC -!- FUNCTION: PROFILIN PREVENTS THE POLYMERIZATION OF ACTIN.
CC -!- OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC ACTIN
CC IN A 1:1 RATIO.
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN COELOMOCYTES IN
CC RESPONSE TO INJURY.
DR EMBL; S42185; S42185.
DR PIR; A44777; A44777.
DR HSSP; P07763; 2ACG.
DR PROSITE; PS00414; PROFILIN.
KW ACTIN-BINDING.
FT INIT MET 0 0 BY SIMILARITY.
SQ SEQUENCE 141 AA; 15151 MW; 104928 CN;

Initial Score = 5 Optimized Score = 5 Significance = 3.38
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
| | |
GIYVNGTKYQFLREEDSKVLGKKKG
80 X 90

6. US-08-121-713B-17 (1-6)
ATPF_THIFE ATP SYNTHASE B CHAIN (EC 3.6.1.34).

ID ATPF THIFE STANDARD; PRT; 159 AA.
AC P41172;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ATP SYNTHASE B CHAIN (EC 3.6.1.34).
GN ATPF.
OS THIOBACILLUS FERROOXIDANS.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC CHEMOLITHOTROPHIC;
OC SULFUR METABOLIZING.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33020;
RM 95047244
RA BROWN L.D., DENNEHY M.E., RAWLINGS D.E.;
RL FEMS MICROBIOL. LETT. 122:19-26(1994).
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -!- SIMILARITY: TO OTHER B SUBUNITS AND ALSO TO B' SUBUNITS.
DR EMBL; M81087; TFUNCAH.
KW HYDROGEN ION TRANSPORT; TRANSMEMBRANE; CF(0).

SQ SEQUENCE 159 AA; 17884 MW; 100442 CN;
Initial Score = 5 Optimized Score = 5 Significance = 3.38
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
|||
I IANAERRGVLEEEAQQKAREEADR
80 90 X 100

7. US-08-121-713B-17 (1-6)
PTH_ECOLI PEPTIDYL-TRNA HYDROLASE (EC 3.1.1.29) (PTH).

ID PTH ECOLI STANDARD; PRT; 194 AA.
AC P23932;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PEPTIDYL-TRNA HYDROLASE (EC 3.1.1.29) (PTH).
GN PTH.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RC STRAIN=K12;
RM 92007806
RA GARCIA-VILLEGAS M.R., DE LA VEGA F.M., GALINDO J.M., SEGURA M.,
BUCKINGHAM R.H., GUARNEROS G.;
RL EMBO J. 10:3549-3555(1991).
CC -!- FUNCTION: THE NATURAL SUBSTRATE FOR THIS ENZYME MAY BE PEPTIDYL-
TRNAS WHICH DROP OFF THE RIBOSOME DURING PROTEIN SYNTHESIS.
CC INVOLVED IN LAMBDA INHIBITION OF HOST PROTEIN SYNTHESIS. PTH
CC ACTIVITY MAY, DIRECTLY OR INDIRECTLY, BE THE TARGET FOR LAMBDA
CC BAR RNA LEADING TO RAP CELL DEATH.
CC -!- CATALYTIC ACTIVITY: N-SUBSTITUTED AMINOACYL-TRNA + H(2)O =
CC N-SUBSTITUTED AMINO ACID + TRNA.
CC -!- SUBUNIT: MONOMER.
CC EMBL; X61941; ECPHGH.
DR PIR; S17791; S17791.
DR PIR; S16753; S16753.
DR ECGENE; EG10785; PTH.
KW HYDROLASE.
FT MUTAGEN 101 101 G->D: PTH(TS) MUTANTS SYNTHESIZE
FT THERMOSENSITIVE PTH AND DIE AT 42 DEGREES
FT FROM A DEFECT IN PROTEIN SYNTHESIS.
FT R->H: RAP MUTANTS DO NOT SUPPORT
FT VEGETATIVE GROWTH OF BACTERIOPHAGE LAMBDA
FT AND DIE UPON TRANSCRIPTION OF LAMBDA DNA
FT BAR SITES.
SQ SEQUENCE 194 AA; 21082 MW; 179056 CN;
Initial Score = 5 Optimized Score = 5 Significance = 3.38
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X

FXREXA
|||
VDLAEERLAPLREAEKFFGYTSRVT
30 40 X 50

8. US-08-121-713B-17 (1-6)
TRJ4_ECOLI TRAJ PROTEIN.

ID TRJ4 ECOLI STANDARD; PRT; 223 AA.
AC P13972;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE TRAJ PROTEIN.
GN TRAJ.
OS ESCHERICHIA COLI.
OG PLASMID R100.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RM 88227859
RA INAMOTO S., YOSHIOKA Y., OHTSUBO E.;
RL J. BACTERIOL. 170:2749-2757(1988).
CC -!- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR POSITIVELY REGULATING THE
CC EXPRESSION OF TRANSFER GENES THAT ARE INVOLVED IN THE CONJUGAL
CC TRANSFER OF DNA BETWEEN BACTERIAL CELLS.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
DR EMBL; M20941; PRTRAJ.
KW OUTER MEMBRANE; CONJUGATION; TRANSCRIPTION REGULATION; ACTIVATOR;
KW PLASMID.
SQ SEQUENCE 223 AA; 25943 MW; 272537 CN;
Initial Score = 5 Optimized Score = 5 Significance = 3.38
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
|||
YSLPVQVATSLREELDAMSLPSSMN
70 X 80

9. US-08-121-713B-17 (1-6)
TPM1_LOCMI TROPOMYOSIN, MUSCLE.

ID TPM1 LOCMI STANDARD; PRT; 283 AA.
AC P31816;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE TROPOMYOSIN, MUSCLE.
OS LOCUSTA MIGRATORIA (MIGRATORY LOCUST).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; ORTHOPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RA KRIEGER J., RAMING K., KNIPPER M., GRAU M., MERTENS S., BREER H.;
RL INSECT BIOCHEM. 20:173-184(1990).

CC -!- FUNCTION: TROPOMYOSIN, IN ASSOCIATION WITH THE TROPONIN COMPLEX,
CC PLAYS A CENTRAL ROLE IN THE CALCIUM DEPENDENT REGULATION OF
CC MUSCLE CONTRACTION.
CC -!- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
DR PIR; A60364; A60364.
DR PROSITE; PS00325; COFILIN TROPOMYOSIN.
DR PROSITE; PS00326; TROPOMYOSIN.
KW MUSCLE PROTEIN; COILED COIL; REPEAT.
SQ SEQUENCE 283 AA; 32439 MW; 334140 CN;

Initial Score = 5 Optimized Score = 5 Significance = 3.38
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
| |
ALENOLKEARFLAEADKKYDEVARK
150 X 160

10. US-08-121-713B-17 (1-6)
TPM2 DROME TROPOMYOSIN II, MUSCLE ISOFORM.

ID TPM2 DROME STANDARD; PRT; 285 AA.
AC P06754;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE TROPOMYOSIN II, MUSCLE ISOFORM.
GN DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RM 87106815
RA HANKE P.D.; STORTI R.V.;
RL GENE 45:211-214(1986).
RN [2]
RP SEQUENCE FROM N.A.
RM 87064486
RA KARLIK C.C.; FYRBERG E.A.;
RL MOL. CELL. BIOL. 6:1965-1973(1986).

CC -!- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -!- ALTERNATIVE PRODUCTS: DROSOPHILA TROPOMYOSIN II GENE CAN PRODUCE
CC FOUR DIFFERENT ISOFORMS BY ALTERNATIVE PRODUCTS: A MUSCLE FORM,
CC A NON-MUSCLE FORM, AND TWO FUSION PROTEINS (33 AND 34).
CC EMBL; M15466; DMTROPII.
DR EMBL; M12840; DMTRO01.
DR EMBL; L00335; DMTRO02.
DR EMBL; L00356; DMTRO03.
DR EMBL; L00357; DMTRO04.
DR EMBL; L00358; DMTRO05.
DR EMBL; L00359; DMTRO06.
DR EMBL; L00360; DMTRO07.
DR EMBL; L00362; DMTRO09.
DR EMBL; M13023; DMTRO13.
DR PIR; A25561; A25561.

DR FLYBASE; FBGN004117; TM2.
DR PROSITE; PS00325; COFILIN TROPOMYOSIN.
DR PROSITE; PS00326; TROPOMYOSIN.
KW MUSCLE PROTEIN; COILED COIL; REPEAT; ALTERNATIVE SPLICING;
KW MULTIGENE FAMILY.
FT CONFLICT 106 114 LGSATKLS -> SASAIQIAA (IN REF. 2).
FT CONFLICT 214 214 S -> A (IN REF. 2).
SQ SEQUENCE 285 AA; 32761 MW; 339904 CN;

Initial Score = 5 Optimized Score = 5 Significance = 3.38
Residue Identity = 50% Matches = 3 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
FXREXA
| |
ALENOLKEARFLAEADKKYDEVARK
150 X 160

11. US-08-121-713B-17 (1-6)
META_ECOLI HOMOSERINE O-SUCCINYLTRANSFERASE (EC 2.3.1.46) (HO

ID META_ECOLI STANDARD; PRT; 309 AA.
AC P07623;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE HOMOSERINE O-SUCCINYLTRANSFERASE (EC 2.3.1.46) (HOMOSERINE O-
GN TRANSUCCINYLAASE).
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RM 89240044
RA COZZONE A.J.;
RL NUCLEIC ACIDS RES. 17:2856-2856(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RM 94089392
RA BLATTNER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.,
RA DANIELS D.L.;
RL NUCLEIC ACIDS RES. 21:5408-5417(1993).
RN [3]
RP SEQUENCE OF 1-69 FROM N.A.
RC STRAIN=K12;
RM 85054633
RA MICHAELI S., MEVARECH M., RON E.Z.;
RL J. BACTERIOL. 160:1158-1162(1984).
CC -!- CATALYTIC ACTIVITY: SUCCINYL-COA + L-HOMOSERINE = COA +
CC O-SUCCINYL-L-HOMOSERINE.
CC -!- PATHWAY: FIRST ENZYME AND ONE OF THE KEY ENZYMES OF METHIONINE
CC BIOSYNTHESIS PATHWAY.
DR EMBL; M10210; ECMETA.
DR EMBL; X14501; ECMETAG.
DR EMBL; U00006; ECUW89.

DR PIR; A05053; XYECM.
 DR ECGENE; EG10581; META.
 KW METHIONINE BIOSYNTHESIS; TRANSFERASE; ACYLTRANSFERASE.
 FT CONFLICT 67 V -> I (IN REF. 2).
 SQ SEQUENCE 309 AA; 35713 MW; 488047 CN;
 Initial Score = 5 Optimized Score = 5 Significance = 3.38
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0
 X X
 FXREXA
 I I I
 RVPDLPVNFLEENVFVMTTGRAS
 10 X 20
 12. US-08-121-713B-17 (1-6)
 STRN_STRGR STRN PROTEIN.
 ID STRN_STRGR STANDARD; PRT; 319 AA.
 AC P29784;
 DT 01-APR-1993 (REL. 25, CREATED)
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
 DE STRN PROTEIN.
 DE STRN PROTEIN.
 GN STRN
 OS STREPTOMYCETES GRISEUS.
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=N2-3-11;
 RM 92092953
 RA PISSOWOTZKI K., MANSOURI K., PIEPERSBERG W.;
 RL MOL. GEN. GENET. 231:113-123(1991).
 CC -!- PATHWAY: STREPTOMYCIN BIOSYNTHESIS.
 CC -!- SIMILARITY: CONTAINS SEQUENCE MOTIFS ALSO CONSERVED IN THE
 CC PUTATIVE CATALYTIC AND/OR SUBSTRATE RECOGNITION DOMAINS OF
 CC AMINOGLYCOSIDE PHOSPHOTRANSFERASES AND EUKARYOTIC PROTEIN
 CC KINASES.
 CC EMBL; X62567; SGSTRELMB.
 DR PIR; S18621; BWSMNG.
 KW STREPTOMYCIN BIOSYNTHESIS.
 SQ SEQUENCE 319 AA; 35679 MW; 470485 CN;
 Initial Score = 5 Optimized Score = 5 Significance = 3.38
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0
 X X
 FXREXA
 I I I
 RVDRCGFLAELEERATRVAGECAAR
 190 200 X 210
 13. US-08-121-713B-17 (1-6)
 I12B_HUMAN INTERLEUKIN-12 BETA CHAIN PRECURSOR (IL-12B) (CYTO
 ID I12B_HUMAN STANDARD; PRT; 328 AA.

P29460;
 DT 01-APR-1993 (REL. 25, CREATED)
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE INTERLEUKIN-12 BETA CHAIN PRECURSOR (IL-12B) (CYTOTOXIC LYMPHOCYTE
 DE MATURATION FACTOR 40 KD SUBUNIT) (CLMF P40) (NK CELL STIMULATORY
 DE FACTOR CHAIN 2) (NKSF2).
 GN I112B OR NKSFP2.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RN SEQUENCE FROM N.A.
 RM 91239523
 RA GUBLER U., CHUA A.O., SCHOENHAUT D.S., DWYER C.M., MCCOMAS W.,
 RA MOTYKA R., NABAVI N., WOLITZKY A.G., QUINN P.M., FAMILLETTI P.C.,
 RA GATELY M.K.;
 RA PROC. NATL. ACAD. SCI. U.S.A. 88:4143-4147(1991).
 RL [2]
 RP SEQUENCE FROM N.A.
 RM 91201875
 RA WOLF S.F., TEMPLE P.A., KOBAYASHI M., YOUNG D., DIGIG M., LOWE L.,
 RA DZIALO R., FITZ L., FERENZ C., HEWICK R.M., KELLEHER K.,
 RA HERRMANN S.H., CLARK S.C., AZZONI L., CHAN S.H., TRINCHIERI G.,
 RA PERUSSIA B.;
 RL J. IMMUNOL. 146:3074-3081(1991).
 RN [3]
 RN SEQUENCE OF 23-45.
 RM 90370873
 RA STERN A.S., PODLASKI F.J., HULMES J.D., PAN Y.C.E., QUINN P.M.,
 RA WOLITZKY A.G., FAMILLETTI P.C., STREMLO D.L., TRUITT T.,
 RA CHIZZONITE R., GATELY M.K.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:6808-6812(1990).
 RN [4]
 RN SIMILARITY TO IL-6 RECEPTOR.
 RM 91300556
 RM GEARING D.P., COSMAN D.;
 RL CELL 66:9-10(1991).
 CC -!- FUNCTION: CYTOKINE THAT CAN ACT AS A GROWTH FACTOR FOR ACTIVATED
 CC T AND NK CELLS, ENHANCE THE LYTIC ACTIVITY OF NK/LYMPHOKINE-
 CC ACTIVATED KILLER CELLS, AND STIMULATE THE PRODUCTION OF IFN-GAMMA
 CC BY RESTING PBMC.
 CC -!- SUBUNIT: DISULFIDE-BONDED HETERODIMER OF 40 KD AND 35 KD SUBUNITS.
 CC NKSF IS ESSENTIALLY A COMPLEX OF CYTOKINE AND SOLUBLE RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC ONE IG-LIKE DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 DR EMBL; M65272; HSCLMF40.
 DR EMBL; M65290; HSNKSP40.
 DR PIR; B39359; B39359.
 DR PIR; B36055; B36055.
 DR MIM; 161561; 11TH EDITION.
 DR PROSITE; PS00241; RECEPTOR_CYTOKINES_1.
 DR PROSITE; PS00340; RECEPTOR_CYTOKINES_2.
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; IMMUNOGLOBULIN FOLD; SIGNAL.
 FT SIGNAL 1 22
 FT CHAIN 23 328 INTERLEUKIN-12 BETA CHAIN.
 FT DOMAIN 23 260 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 261 279 POTENTIAL.

FT DOMAIN 280 328 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 43 97 IG-LIKE DOMAIN (BY SIMILARITY).
 FT DISULFID 50 90 POTENTIAL.
 FT CARBOHYD 135 135 POTENTIAL.
 FT CARBOHYD 222 222 POTENTIAL.
 FT CARBOHYD 303 303 K -> N (IN REF. 2).
 FT CONFLICT 239 239
 SQ SEQUENCE 328 AA; 37169 MW; 606556 CN;
 Initial Score = 5 Optimized Score = 5 Significance = 3.38
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXREXA
 | | |
 KDOKEPKNKTFLRCEAKNYSGRFTCW
 120 130 X 140

14. US-08-121-713B-17 (1-6)
 YHA9 YEAST HYPOTHETICAL 38.0 KD PROTEIN IN PRPS4-STE20 INTERG

ID YHA9 YEAST STANDARD; PRT; 330 AA.
 AC P38749;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 38.0 KD PROTEIN IN PRPS4-STE20 INTERGENIC REGION.
 GN YHL009C.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RM 94378003
 RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
 RA DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,
 RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
 RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
 RA NHAN M., RIFKIN L., RILES L., ST.PETER H., TREVASKIS E., VAUGHAN K.,
 RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
 RA VAUDIN M.;
 RL SCIENCE 265:2077-2082(1994).
 CC -!- SIMILARITY: SOME TO BZIP PROTEINS.
 DR EMBL; U11581; SCHL5018.
 DR PIR; S46819; S46819.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 330 AA; 37955 MW; 539642 CN;
 Initial Score = 5 Optimized Score = 5 Significance = 3.38
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXREXA
 | | |
 GPAYPRSLIDFLIVEEATPLNE
 320 X 330

15. US-08-121-713B-17 (1-6)
 BCHX_RHOCA CHLOROPHYLLIDE REDUCTASE 35.5 KD CHAIN (EC 1.3.1.1.-

ID BCHX_RHOCA STANDARD; PRT; 333 AA.
 AC P26177;
 DT 01-MAY-1992 (REL. 22, CREATED)
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE CHLOROPHYLLIDE REDUCTASE 35.5 KD CHAIN (EC 1.3.1.1.-) (CHLORIN
 REDUCTASE).
 GN BCHX.
 OS RHODOBACTER CAPSULATUS (RHODOPSEUDOMONAS CAPSULATA).
 OC PROKARYOTA; GRACILICUTES; ANOXYPHOTOBACTERIA; PURPLE BACTERIA;
 OC RHODOSPIRILLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SB1003;
 RM 93224464
 RA BURKE D.H., ALBERTI M., HEARST J.E.;
 RL J. BACTERIOL. 175:2407-2413(1993).
 CC -!- FUNCTION: CONVERTS CHLOROPHYLLS (CHL) INTO BACTERIOCHLOROPHYLLS
 (BCHL) BY REDUCING RING B OF THE TETRAPYRROLE.
 CC -!- PATHWAY: BACTERIOCHLOROPHYLL BIOSYNTHESIS.
 CC -!- SIMILARITY: BELONGS TO THE NIFH/FRXC FAMILY.
 DR EMBL; Z11165; RCPHSYNG.
 DR PIR; S17823; S17823.
 DR PIR; B49850; B49850.
 DR HSSP; P00459; INIP.
 DR PROSITE; PS00692; NIFH FRXC 2.
 DR PROSITE; PS00746; NIFH FRXC 1.
 KW PHOTOSYNTHESIS; CHLOROPHYLL BIOSYNTHESIS; ATP-BINDING; IRON-SULFUR;
 KW 4FE-4S.
 FT NP BIND 42 49 ATP (POTENTIAL).
 FT METAL 130 130 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 165 165 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 333 AA; 35573 MW; 536770 CN;
 Initial Score = 5 Optimized Score = 5 Significance = 3.38
 Residue Identity = 50% Matches = 3 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXREXA
 | | |
 DAPNLKGFDAIRLEEAEEPTLEIPE
 10 X 20

maryh@stic

stdin

NeWSprinter20

Fri May 19 10:57:16 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5

> O <
O| | O IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq22asq.res made by on Fri 19 May 95 8:41:38-PDT.

Query sequence being compared: US-08-121-713B-22 (1-7)
Number of sequences searched: 53402
Number of scores above cutoff: 4031

Results of the initial comparison of US-08-121-713B-22 (1-7) with:
Data bank : A-GeneSeq 18, all entries

100000-

-

N

U50000-

M

B

E

R

-

O

F10000-

S

E 5000-

Q

D

E

N

C

E

S 1000-

-

500-

-

100-

-

50-

-

10-

-

SDQ 22

5-

-

-

-

-

0

||

SCORE 0|

STDEV 0

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

||

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	6
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	1	1	1.78

Times: CPU 00:00:32.97 Total Elapsed 00:00:33.00

Number of residues: 6354270

Number of sequences searched: 53402

Number of scores above cutoff: 4031

Cut-off raised to 3.

Cut-off raised to 4.

Cut-off raised to 5.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

1761 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig.	Frame
1. R44929	T. niveum Cyclosporin synthet	15281	7	7	3.37	0
2. R11510	Ryanodine receptor deduced fr	5072	7	7	3.37	0
3. R25450	MH mutant porcine ryanodine r	5035	7	7	3.37	0
4. R10834	Rianodin receptor.	4987	7	7	3.37	0
5. R47861	Alpha 2-Macroglobulin/IDL-rec	4544	7	7	3.37	0
6. R52971	Product of the cDNA encoding	3969	7	7	3.37	0
7. R38470	ALL-1 protein.	3910	7	7	3.37	0
8. R06996	Protein characteristic of hog	3898	7	7	3.37	0
9. R10473	Hog cholera virus genome prod	3898	7	7	3.37	0
10. R13895	ACV synthetase.	3778	7	7	3.37	0

11. R13753 ACVS.
 12. R10145 Cephalosporin antibiotic bios
 13. R13896 ACV synthetase.
 14. R30290 Human Duchenne muscular dyst
 15. R30373 Sequence encoded by human mus
 16. R5041 Filamentous haemagglutinin A.
 17. R40227 ACVS.
 18. R34712 Bacillus subtilis srfA operon
 19. R34713 Bacillus subtilis srfA operon
 20. R43662 DNM1-S275/90 (BCACC V92042111
 21. R31770 Deduced sequence encoded by m
 22. R31771 Deduced sequence encoded by b
 23. R3285 Sequence of clone HIV-2 SBL/I
 24. R39926 GAP protein of Ira2.
 25. R40841 Translation of TEV large ORF.
 26. R33539 NANBH virus strain HC-J8 prot
 27. R33214 NANBH virus strain HC-J8 prot
 28. R33538 NANBH virus strain HC-J6 prot
 29. R3284 Sequence of clone HIV-2 SBL/I
 30. R35207 Hepatitis C virus protein.
 31. R34099 NANBHV E1/E2 protein.
 32. R21519 Compiled HCV sequence.
 33. R22154 NANBH Hutch c59 isolate genom
 34. R31621 Hepatitis C virus (HCV) poly
 35. R34468 Encoded by full-length Hepati
 36. R40119 HCV genomic amino acid sequen
 37. R40120 HCV genomic amino acid sequen
 38. R20091 Non-A, non-B viral genome pro
 39. R20111 Non-A, non-B viral genome pro
 40. R30616 Polypeptide coded by Korean H
 41. R34580 Human hepatitis C virus gene
 42. R33417 Blood transmissible NANBHV pr
 43. R44432 eYvA region polypeptide modul
 44. R08124 Hepatitis C virus putative po
 45. R39923 GAP protein Iral.

1. US-08-121-713B-22 (1-7)
 R44929 T. niveum Cyclosporin synthetase.

ID R44929 standard; Protein; 15281 AA.
 AC R44929;
 DT 08-JUL-1994 (first entry)
 DE T. niveum Cyclosporin synthetase.
 KW Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum;
 KW T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase.
 OS Tolypocladium niveum.
 PN EF-578616-A.
 PD 12-JAN-1994.
 PF 05-JUL-1993; 810474.
 PR 03-JUL-1992; AT-001403.
 PR 28-MAR-1993; AT-000437.
 PR 29-APR-1993; CH-001310.
 PR 04-MAY-1993; CH-001375.
 PA (SANO) SANDOZ LTD.
 PA (SANO) SANDOZ PATENT GMBH.
 PA (SANO) SANDOZ-ERINDUNGEN VERW GES MEH.
 PI Leither E, Schneider E, Schoergendorfer K, Weber G;
 DR WPI; 94-010432/02.
 DR N-PSDB; Q54386.

PT Isolated DNA sequence - which codes for enzyme having cyclosporin
 PT synthetase like activity
 PS Claim 1; Page 41-84; 93pp; English.
 CC This sequence represents an enzyme which has cyclosporin synthetase-
 CC like activity. This sequence was isolated from Tolypocladium niveum
 CC (formerly known as T. inflatum GAMS). This enzyme catalyses the
 CC peptide biosynthesis of cyclosporins and structurally related
 CC molecules. This sequence may be used for the production of
 CC cyclosporin by transforming a vector containing this sequence in
 CC to a recombinant host. This allows effective production of anti-
 CC biotic cyclosporin or its derivatives.
 SQ Sequence 15281 AA;
 SQ 1431A; 1009R; 523N; 950D; 0 B; 101C; 7850; 972E; 0 Z; 946G; 379H;
 SQ 851I; 156L; 478K; 327M; 575F; 731P; 106S; 892T; 138W; 337Y; 123V;
 Initial Score = 7 Optimized Score = 7 Significance = 3.37
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X X
 DXVXXX
 |
 ASQPSLDVHDLVQLAEDAGFVEVSW
 2390 X 2400 2410

2. US-08-121-713B-22 (1-7)
 R11510 Ryanodine receptor deduced from cDNA clone.

ID R11510 standard; Protein; 5072 AA.
 AC R11510;
 DT 17-JUN-1991 (first entry)
 DE Ryanodine receptor deduced from cDNA clone.
 DE Malignant hyperthermia; hypermetabolic syndrome; inhalation;
 KW anaesthetics; probe; calcium release channel; sarcoplasmic;
 KW reticulum.

OS Homo sapiens.
 FH Key
 FT Peptide 1629..1632 Location/Qualifiers
 FT Peptide 1742..1748
 FT Peptide 3119..3130
 FT Peptide 3196..3210
 FT Active-site 3981..3985
 FT /label= phosphorylation site
 FT Active-site 4315..4318
 FT /label= phosphorylation site
 PN W09104328-A.
 PD 04-APR-1991.
 PF 21-SEP-1990; CA0312.
 PR 25-SEP-1989; US-612726.
 PA (HSCR-) HSC R & D Partnership.
 PA (UTOR) Univ. Toronto Innova.
 PA (TORO-) Toronto Hospital.
 PI Worton RG, MacLennan DH, Britt BA.
 DR WPI; 91-117517/16.
 DR Q-PSDB; Q11415.

PT Purified DNA specific for human ryanodine receptor - useful for
 PT diagnosis of malignant melanoma.
 PS Claim 11; Fig 2; 49pp; English.
 CC The sequence was obtd. from several overlapping clones isolated

CC from a human skeletal muscle cDNA library in lambda gt10. The
 CC protein encodes the human ryanodine receptor (HRR), a calcium
 CC release channel which spans the gap between the transverse tubule
 CC and the sarcoplasmic reticulum (SR) in muscle. The sustained muscle
 CC contraction in malignant hyperthermia (MH) may be caused by the
 CC release of calcium into the muscle cell cytoplasm from the SR. This
 CC is due to a defect in the gene encoding HRR. Sequences from the
 CC gene can be used for diagnosis of MH using RFLP analysis. The
 CC protein sequence would give rise to several transmembrane passages
 CC in the C-terminal fifth of the molecule; the bulk would be hydro-
 CC philic. This matches well with the structure of known ryanodine
 CC receptors. The peptides in the feature table were determined from
 CC the purified receptor protein.

SQ Sequence 5072 AA;
 SQ 387A; 295R; 174N; 252D; 0 B; 100C; 207Q; 477E; 0 Z; 362G; 134H;
 SQ 207I; 561L; 230K; 149W; 209P; 274P; 297S; 232T; 63 W; 142Y; 320V;

Initial Score = 7 Optimized Score = 7 Significance = 3.37
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X
 DXVXXXX
 |
 PGQGGRISTDLVIGICVLDTATGLMTF
 1510 1520

3. US-08-121-713B-22 (1-7)
 R25450 MH mutant porcine ryanodine receptor.

ID R25450 standard; Protein; 5035 AA.
 AC R25450;
 DT 08-JAN-1993 (first entry)
 DE MH mutant porcine ryanodine receptor.
 KW MH; RYR1; calcium release channel; sarcoplasmic reticulum;
 KW transverse tubule; polymorphism; beta strand.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc difference 615
 FT /label= MH mutation
 PN WO9211387-A.
 PD 09-JUL-1992.
 PF 20-DEC-1991; CA0457.
 PR 21-DEC-1990; GB-027869.
 PR 20-MAY-1991; GB-010865.
 PR 09-SEP-1991; GB-019250.
 PA (YUGU-) UNIV GUELPH.
 PA (UTOR-) UNIV TORONTO INNOVATIONS FOUND.
 PI MacLennan DH, O'Brien PJ;
 DR WPI; 92-250106/30.
 DR N-PSDB; Q25975.
 PT Purified DNA mol. for diagnosis of porcine malignant hyperthermia
 PT - comprises DNA sequence encoding normal or mutant ryanodine
 PT receptor with specified endonuclease restriction map
 PS Disclosure; Fig 2; 96pp; English.
 CC The sequence given is encoded by the mutant pig ryanodine receptor
 CC (RYR1) cDNA. A polymorphic change at position 1972 of the cDNA causes
 CC the mutation from Arg to Cys and this is thought to be the molecular
 CC basis of porcine malignant hyperthermia (MH). This mutation lies

CC within the region of RYR1 that is concerned with the binding of
 CC regulators of Ca2+ release channel gating. Analysis of surrounding
 CC sequences suggests that this mutation lies within a beta strand domain
 CC comprising roughly of amino acids 520 to 830. RYR1 is the calcium
 CC release channel of the sarcoplasmic reticulum and is a large protein
 CC which spans the gap between the transverse tubule and the sarcoplasmic
 CC reticulum. The channel is activated by ATP, calcium, caffeine, and
 CC micro-molar ryanodine. It is inhibited by ruthenium red, tetracaine,
 CC calmodulin, high Mg2+ and ryanodine.

SQ Sequence 5035 AA;
 SQ 381A; 297B; 171N; 248D; 0 B; 100C; 198Q; 486E; 0 Z; 371G; 132H;
 SQ 202I; 568L; 227K; 143M; 206F; 271P; 286S; 227T; 64 W; 143Y; 314V;

Initial Score = 7 Optimized Score = 7 Significance = 3.37
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X
 DXVXXXX
 |
 AIIRSLVPLDILVIGIISLPQLQIPTLKG
 2460 2470 X 2480

4. US-08-121-713B-22 (1-7)
 R10834 Rianodin receptor.

ID R10834 standard; Protein; 4987 AA.
 AC R10834;
 DT 26-APR-1991 (first entry)
 DE Rianodin receptor.
 KW Rianodin receptor gene; calcium release modulator; tranquilliser;
 KW antagonist.
 OS Oryctolagus cuniculus.
 PN J03011098-A.
 PD 18-JAN-1991.
 PF 07-JUN-1989; 144569.
 PR 07-JUN-1989; JP-144569.
 PA (MITU) Mitsubishi Kasei Corp.
 DR WPI; 91-062003/09.
 DR N-PSDB; Q10613.
 PT New rianodin receptor, genes encoding it and its prepn. - useful as
 PT calcium release modulator for tranquillisers and for assaying
 PT calcium antagonists.
 PS Disclosure; Fig 1; 18pp; Japanese.
 CC RNA contg. poly(A) was prepared from rabbit skeletal muscle
 CC endoplasmic reticulum. From the obtd. poly(A) mRNA, a cDNA bank
 CC corresp. to it was prepared by random primer method, oligo (dT)
 CC primer method, and primer extension method. A cDNA was obtained by
 CC screening with a DNA probe (see Q10614-15). By introducing the obtd.
 CC cDNA into an expression vector, vector pRES7 was formed.
 CC The product is said to be involved in calcium release from
 CC sarcoplasmic reticulum which triggers contraction of skeletal muscle.
 CC Therefore, the receptor is useful as tranquilliser and assay series for
 CC screening of calcium antagonist.
 SQ Sequence 4987 AA;
 SQ 390A; 296R; 170N; 249D; 0 B; 101C; 198Q; 483E; 0 Z; 365G; 131H;
 SQ 192I; 551L; 224K; 145W; 206F; 285P; 278S; 221I; 65 W; 142I; 315V;

Initial Score = 7 Optimized Score = 7 Significance = 3.37

Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXX

PGQGRISHTDLVIGCLVLAATGLMTF
1510 X 1520

5. US-08-121-713B-22 (1-7)
R47861 Alpha 2-Macroglobulin/LDL-receptor related protein

ID R47861 standard; protein; 4544 AA.
AC R47861;
DT 20-JUL-1994 (first entry)
DE Alpha 2-Macroglobulin/LDL-receptor related protein.
KW alpha-2 macroglobulin; Low Density Lipoprotein; LDL; receptor family;
KW LDL receptor related protein; LRP; small rhinovirus receptor; deriv;
KW Minor Rhinovirus; alpha2MR/LRP.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc difference 211..260
FT /note= "50 residues not shown in SEQ.ID.No.4"
FT Misc difference 1990
FT /note= "Residue not shown in SEQ.ID.No.4"
FT Misc difference 3050
FT /note= "Residue not shown in SEQ.ID.No.4"
FN W09401553-A.
PD 20-JAN-1994.
PF 05-JUL-1993; E01728.
PR 08-JUL-1992; DE-222385.
PR 22-AUG-1992; DE-227892.
PR 19-FEB-1993; DE-305063.
PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.
PI Blaas D, Gruenberger M, Hofer F, Huettlinger M, Kerjaeschki D;
PI Kowalski H, Kuechler E, Machat H;
DR WPI; 94-035060/04.
PT New peptide derivs. of receptor for rhinovirus - of the small
PT receptor gp., and derived DNA, transformed cells and antibodies,
PT used e.g. to treat or prevent rhinovirus infection
PS Claim 5; Fig 2; 76pp; German.
CC Functional derivatives of members of the Minor Rhinovirus Receptor
CC group are claimed. The alpha-2 Macroglobulin/LDL-receptor related
CC protein of sequence R47861 (Herz et al. EMBO J. 7:4119-4127 (1988))
CC is a preferred parent receptor. The derivs, which are preferably
CC soluble, extracellular forms of the native receptors, are useful
CC for treating and preventing viral (esp. rhinoviral) infections.
CC N.B. the SEQ.ID. listing includes a sequence (no.4) which differs
CC from the alpha2-MR/LRP sequence as indicated in the Features Table.
SQ Sequence 4544 AA;
SQ 249A; 277R; 249N; 393D; 0 B; 331C; 164Q; 235E; 0 Z; 376G; 129H;
SQ 196I; 305L; 173K; 72 M; 133F; 231P; 308S; 259T; 88 W; 131Y; 245V;

Initial Score = 7 Optimized Score = 7 Significance = 3.37
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXX

KIVFPHGITDLVLSRLVWADAYLDYI
370 X 380

6. US-08-121-713B-22 (1-7)
R52971 Product of the cDNA encoding htrx.

ID R52971 standard; Protein; 3969 AA.
AC R52971;
DT 27-SEP-1994 (first entry)
DE Product of the cDNA encoding htrx.
KW Human; trithorax gene; L01986; diagnosis; treatment;
KW immunodeficiency; developmental abnormalities; inherited diseases;
KW cancer; acute lymphocytic leukaemia; myelomonocytic leukaemia.
OS Homo sapiens.
PN W09407502-A.
PD 14-APR-1994. U09087.
PF 24-SEP-1993;
PR 30-SEP-1992; US-954112.
PR 13-MAY-1993; US-061376.
PA (SALK) SALK INST BIOLOGICAL STUDIES.
PI Djabaki M, Evans GA, Parry P, Selleri L;
DR WPI; 94-135206/16.
PT Nucleic acid encoding a human tri:thorax protein - used to
PT develop agents for diagnosis and treatment of diseases associated
PT with disruption of chromosome II at q23
PS Disclosure; Page 43-54; 68pp; English.
CC In the course of the construction of a physical map of human
CC chromosome region 11q23, a region contg. the t(4;11) translocation
CC breakpoint was cloned. The cloned DNA encoded a protein homologous
CC to the trithorax gene prod. of Drosophila. The gene may be used for
CC the diagnosis and treatment of immunodeficiency states.
CC developmental abnormalities, inherited diseases or cancers, e.g.
CC acute lymphocytic leukaemia or acute myelomonocytic leukaemia.
CC See also R52972-7.
SQ Sequence 3969 AA;
SQ 218A; 223R; 159N; 182D; 0 B; 82 C; 184Q; 250E; 2 Z; 249G; 86 H;
SQ 143I; 286L; 299K; 64 M; 103F; 352P; 543S; 245T; 22 W; 55 Y; 203V;
SQ 19 Others;

Initial Score = 7 Optimized Score = 7 Significance = 3.37
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXX

SSEMKOSSASDIVSKSSSLKGEKTKVL
2290 X 2300 2310

7. US-08-121-713B-22 (1-7)
R38470 ALL-1 protein.

ID R38470 standard; Protein; 3910 AA.
AC R38470;
DT 08-NOV-1993 (first entry)
DE ALL-1 protein.
KW Acute lymphoblastic leukemia gene; ALL-1; chromosome 11; treatment;

KW translocation breakpoint mapping; chromosomal abnormality; diagnosis;
 KW human; acute lymphocytic; myelomonocytic; monocytic; myelogenous;
 KW leukemia; Drosophila; trithorax; homology region; zinc finger domain;
 KW cysteine-rich.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 1021..1221
 FT /note= "Region of homology to Drosophila trithorax"
 FT Region 1462..1570
 FT /note= "Region of homology to Drosophila trithorax"
 FT Region 3348..3562
 FT /note= "Region of homology to Drosophila trithorax"
 PN W09312136-A.
 PD 24-JUN-1993.
 PF 09-DEC-1992; U10930.
 PR 11-DEC-1991; US-805093.
 PR 27-MAY-1992; US-888839.
 PR 30-OCT-1992; US-971094.
 PA (UJJE-) UNIV JEFFERSON THOMAS.
 PI Canaan E, Croce CM;
 DR WPI; 93-214090/26.
 DR N-PSDB; Q43526.
 PT Detection and treatment of acute leukaemia(s) - using prods.
 PT derived from oligo:nucleotide sequences within the ALL-1 gene of
 PT chromosome 11
 PS Disclosure; Page 29-50; 90pp; English.
 CC This sequence is encoded by the acute lymphoblastic leukemia (ALL-1)
 CC gene of chromosome 11. The ALL-1 gene was isolated by translocation
 CC breakpoint mapping. Fragments of the ALL-1 cDNA may be used to
 CC identify chromosomal abnormalities within the ALL-1 gene. These
 CC leukemias such as acute lymphocytic, myelomonocytic, monocytic and
 CC myelogenous leukemia. ALL-1 protein shows three regions of homology
 CC to the Drosophila trithorax protein. These regions show 64%, 66% and
 CC 82% similarity respectively, to the Drosophila gene. The third region
 CC of homology constitutes the extreme C-terminus of the two proteins,
 CC both proteins end in an identical sequence. The first homology region
 CC is cysteine-rich and contains sequence motifs analogous to four zinc
 CC finger domains (3-6) within the trithorax gene. The second region of
 CC homology is also cysteine-rich and corresponds to zinc fingers 7 and 8
 CC of the Drosophila gene. The multiple conserved cysteines and
 CC histidines at the 3' end of the motifs allow two or three arrangements
 CC of the putative fingers. The structure of these cysteine-rich domains
 CC appears to be unique to the trithorax and ALL-1 genes.
 SQ Sequence 3910 AA;
 SQ 214A; 222R; 159N; 181D; 0 B; 79 C; 180Q; 253E; 0 Z; 230G; 83 H;
 SQ 149I; 275L; 309K; 63 M; 100F; 345P; 542S; 230T; 20 W; 53 Y; 207V;
 Initial Score = 7 Optimized Score = 7 Significance = 3.37
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

SSEMKQSSADLVSKSSIKGEKTKVL
 2230 2240 X 2250

8. US-08-121-713B-22 (1-7)

R06996 Protein characteristic of hog cholera virus (HCV).
 ID R06996 standard; protein; 3898 AA.
 AC R06996;
 DT 16-JAN-1991 (first entry)
 DE Protein characteristic of hog cholera virus (HCV).
 KW Vaccine; parvovirus; swine influenza virus.
 OS Hog cholera virus.
 PN EP-389034-A.
 PD 26-SEP-1990.
 PF 12-MAR-1990; 200573.
 PR 19-MAR-1989; EP-104921.
 PR 12-MAR-1990; EP-200573.
 PA (ALKU) AKZO NV.
 PI Meyers G, Rumenapf T, Thiel HJ;
 DR WPI; 90-291979/39.
 DR N-PSDB; Q06001.
 PT New hog cholera virus vaccine and diagnostic - comprises nucleic
 PT acid sequence of poly-peptide characteristic of hog cholera virus
 PS Claim 2; Fig 2; 29pp; English.
 CC Gene product may be used to provide a vaccine and Abs for diagnosis
 CC of hog cholera viral infection in pigs.
 SQ Sequence 3898 AA;
 SQ 243a; 207B; 154N; 184D; 0 B; 72 C; 120Q; 253E; 0 Z; 286G; 82 H;
 SQ 211I; 384L; 296K; 79 M; 124F; 168P; 188S; 303T; 63 W; 165Y; 316V;
 Initial Score = 7 Optimized Score = 7 Significance = 3.37
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

ISDHIKQATDILVYIINRPQFGDT
 2440 2450 X 2460

9. US-08-121-713B-22 (1-7)
 R10473 Hog cholera virus genome product.
 ID R10473 standard; Protein; 3898 AA.
 AC R10473;
 DT 09-APR-1991 (first entry)
 DE Hog cholera virus genome product.
 KW Swine fever virus; HCV; pestivirus; border disease virus; BDV;
 KW Bovine viral diarrhoea virus; BVDV; Togaviridae; ss.
 OS Hog Cholera Virus.
 PN W09100352-A.
 PD 10-JAN-1991.
 PF 29-JUN-1990; N10092.
 PR 29-JUN-1989; NL-001651.
 PA (DIER-) CENT DIERGENESKUND.
 PI Moorman R, Wensvoort G;
 DR WPI; 91-036746/05.
 DR P-PSDB; Q10353.
 PT Protection of animals against microbial infection - using
 PT nucleotide sequences related to the microbe and a stop transfer
 PT region
 PS Claim 15; Fig 1; 33pp; English.
 CC The genome sequence and its encoded product are sufficiently similar

CC to other pestviruses of the family Togaviridae, that they may be
CC used in vaccination and diagnosis of pestivirus infection.
SQ Sequence 3898 AA;
SQ 248A; 198R; 159N; 179D; 0 B; 73 C; 122Q; 258E; 0 Z; 287G; 80 H;
SQ 210I; 381L; 290K; 94 M; 124F; 172P; 185S; 306T; 63 W; 165Y; 304V;
Initial Score = 7 Optimized Score = 7 Significance = 3.37
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X
DXVXXXX
I I
IADHVQAAATDLVVYIINRPQFPQD
2440 2450 X 2460
10. US-08-121-713B-22 (1-7)
R13895 ACV synthetase.
ID R13895 standard; Protein; 3778 AA.
AC R13895;
DT 22-NOV-1991 (first entry)
DE ACV synthetase.
KW Beta lactam antibiotics; penicillin.
OS Penicillin chrysogenum.
FH Key Location/Qualifiers
FT Domain 301..1068
FT /label= I
FT /function= activation of amino acid substrate
FT Region 374..423
FT /label= subdomain
FT Region 474..501
FT /label= subdomain
FT Region 655..699
FT /label= subdomain
FT Region 725..754
FT /label= subdomain
FT Domain 1392..2154
FT /label= II
FT /function= activation of amino acid substrate
FT Region 1470..1518
FT /label= subdomain
FT Region 1564..1590
FT /label= subdomain
FT Region 1745..1789
FT /label= subdomain
FT Region 1817..1846
FT /label= subdomain
FT Domain 2474..3295
FT /label= III
FT /function= activation of amino acid substrate
FT Region 2554..2603
FT /label= subdomain
FT Region 2647..2673
FT /label= subdomain
FT Region 2827..2871
FT /label= subdomain
FT Region 2899..2928
FT /label= subdomain

FT Domain 3560..3647
FT /label= IV
FT /function= thioesterase
PN EP-445868-A.
PD 11-SEP-1991.
PF 27-FEB-1991; 200423.
PR 28-FEB-1990; EP-200475.
PR 28-FEB-1990; EP-200488.
PR 02-JUL-1990; EP-201768.
PR 03-OCT-1990; EP-202628.
PR 27-FEB-1991; EP-200423.
PA (KONN) GIST-BROCADES NV.
PI Veenstra AE, Martin JF, Garcia BD, Guttierrez S, Barredo JL;
PI Montenegro PE, Von Doehren H, Palissa H, Van Liempt H;
DR WPI; 91-268735/37.
DR N-PSDB; Q13607.
PT DNA encoding amino:adipyl-cysteinyI-valine synthetase - used for
PT prodn. of the enzyme or enhanced prodn. of new or known
PT beta-lactam antibiotic cpds.
PS Claim 1; Page 20; 54pp; English.
CC The isolation of DNA fragments encoding the ACV synthetase is
CC described in EP-357119. Cosmid HM193 contains one such fragment
CC which was sequenced using the Sequenase system 2.0. The protein
CC sequence was deduced from the DNA. Three distinct regions of
CC homology have been identified, domains I, II and III. Within
CC these domains several even more conserved elements can be
CC distinguished. Since the enzyme synthesises a tripeptide, which
CC most probably requires the activation of three amino acids, a
CC role for these domains in the activation reactions seems likely.
CC A fourth domain is thought to act as a thioesterase.
CC The gene can be used to express the synthetase enzyme which can
CC be used for the prodn of new beta-lactam antibiotics.
CC See also R13896.
SQ Sequence 3778 AA;
SQ 266A; 245R; 158N; 219D; 0 B; 44 C; 170Q; 260E; 0 Z; 228G; 103H;
SQ 204I; 408L; 140K; 65 M; 146F; 169P; 289S; 209T; 36 W; 137Y; 282V;
Initial Score = 7 Optimized Score = 7 Significance = 3.37
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X
DXVXXXX
I I
GRNSRLYKTDLVRWIPGSGGEVEYIG
1800 1810 X 1820
11. US-08-121-713B-22 (1-7)
R13753 ACVS.
ID R13753 standard; Protein; 3768 AA.
AC R13753;
DT 14-NOV-1991 (first entry)
DE ACVS.
KW Delta-(L-alpha-amino:adipyl)-L-cysteinyI-D-valine synthetase;
KW beta-lactam antibiotics.
OS Penicillium chrysogenum.
FH Key Location/Qualifiers
FT Domain 301..1068

FT /label= domain I 1392..2154
FT Domain 1392..2154
FT /label= domain II
FT Domain 2474..3295
FT /label= domain III
FT EP-444758-A.
PN 04-SEP-1991.
PD 27-FEB-1991. 200422.
PR 28-FEB-1990; EP-200475.
PR 28-FEB-1990; EP-200488.
PR 02-JUL-1990; EP-201768.
PR 03-OCT-1990; EP-202628.
PR 27-FEB-1991; EP-200422.
PA (KONN) GIST-BROCADES NV.
PI Veenstra AE, Martin JF, Garcia BD, Gutierrez S, Barredo JL;
PI Montenegro Prieto E, Von Doehren H, Palissa H, Van Liempt H.
DR WPI; 91-261325/36.
DR N-PSDB; Q13947.
FT Mutant delta-(L-alpha-aminoacyl)-L-cysteiny-D-valine
FT synthetase - used in prodn. of beta-lactam antibiotics.
PS Disclosure; Page 19; 56pp; English.
CC The amino acid sequence codes for delta- (L-alpha-aminoacyl)-L-
cysteiny-D-valine synthetase (ACVS). The prods. may be used for
CC the enhanced expression (in vivo and in vitro) of mutant enzymes
CC and fermentable or known and new beta- lactam antibiotics and their
CC precursors, partic. antibiotics of the penam and cephem classes.
CC See also R13754-R13756.
SQ Sequence 3768 AA;
SQ 266A; 245R; 158N; 218D; 0 B; 44 C; 169Q; 260E; 0 Z; 227G; 104H;
SQ 204I; 407L; 138K; 64 M; 146P; 167P; 288S; 208T; 36 W; 137Y; 282V;
SQ 204I; 407L; 138K; 64 M; 146P; 167P; 288S; 208T; 36 W; 137Y; 282V;
Initial Score = 7 Optimized Score = 3.37
Residue Identity = 28% Matches = 7 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X
DXVXXXX
GRNSRLYKTDGLVRWIPGSSGEVEYL
1790 1800 X 1810
12. US-08-121-713B-22 (1-7)
R10145 Cephalosporin antibiotic biosynthetic enzyme #1.
ID R10145 standard; Protein; 3722 AA.
AC R10145;
DE Cephalosporin antibiotic biosynthetic enzyme #1.
KW Cephalosporin; antibiotic;
KW S-(L-alpha-aminoacyl)-L-cysteiny-D-; valine synthetase;
KW isopenicillin N synthetase; isopenicillin N epimerase;
KW deacetoxycephalosporin C synthetase; beta-lactamase;
KW deacetoxycephalosporin C hydroxylase.
OS Lysobacter lactamgenus.
FT J02291274-A.
PN 03-DEC-1990.
PD 10-JAN-1990; 003762.
PR 01-FEB-1989; JP-024710.
PR 10-JAN-1990; JP-003762.

PA (TAKE) TAKEDA CHEMICAL IND KK.
DR WPI; 91-018854/03.
DR N-PSDB; Q10190.
PT Prepn. of cephalosporin series antibiotics - comprises culturing
PT transformant of microbe transformed by plasmid contg. new DNA
PT fragment
PS Disclosure; Fig 13; 67pp; Japanese.
CC This protein is encoded by ORF1 of the 23666bp sequence
CC isolated from L.lactamgenus and comprising the genes for the
CC cephalosporin biosynthetic enzymes listed in the KEYWORDS. Plasmids
CC containing at least one of ORF's 1-9 can be used to transform
CC microbes, such as bacteria or yeast.
CC See also Q10191-2.
SQ Sequence 3722 AA;
SQ 454A; 307R; 101N; 266D; 0 B; 36 C; 138Q; 233E; 0 Z; 267G; 104H;
SQ 157I; 418L; 84 K; 51 M; 153F; 171P; 218S; 151T; 45 W; 120Y; 248V;
Initial Score = 7 Optimized Score = 3.37
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X
DXVXXXX
GNNARLYKTDGLVRWIPNGELEYLGRN
1750 X 1760 1770
13. US-08-121-713B-22 (1-7)
R13896 ACV synthetase.
ID R13896 standard; Protein; 3712 AA.
AC R13896;
DT 22-NOV-1991 (first entry)
DE ACV synthetase.
KW Beta lactam antibiotics; penicillin.
OS Acetomonium chrysogenum.
FH Key Location/Qualifiers
FT Domain 301..1068
FT /label= I
FT /function= activation of amino acid substrate
FT Region 374..423
FT /label= subdomain
FT Region 474..501
FT /label= subdomain
FT Region 655..699
FT /label= subdomain
FT Region 725..754
FT /label= subdomain
FT Domain 1392..2154
FT /label= II
FT /function= activation of amino acid substrate
FT Region 1470..1518
FT /label= subdomain
FT Region 1564..1590
FT /label= subdomain
FT Region 1745..1789
FT /label= subdomain
FT Region 1817..1846
FT /label= subdomain

FT Domain 2474...3295
 FT /label= III
 FT /function= activation of amino acid substrate
 FT Region 2554...2603
 FT /label= subdomain
 FT Region 2647...2673
 FT /label= subdomain
 FT Region 2827...2871
 FT /label= subdomain
 FT Region 2899...2928
 FT /label= subdomain
 FT Domain 3560...3647
 FT /label= IV
 FT /function= thioesterase
 FT EP-445868-A.
 PD 11-SEP-1991.
 PF 27-FEB-1991; 200423.
 PR 28-FEB-1990; EP-200475.
 PR 28-FEB-1990; EP-200488.
 PR 02-JUL-1990; EP-201768.
 PR 03-OCT-1990; EP-202628.
 PR 27-FEB-1991; EP-200423.
 PA (KONN) GIST-BROCADES NV.
 PI Veenstra AE, Martin JF, Garcia BD, Guttierrez S, Barredo JL;
 PI Montenegro PE, Von Doehren H, Falissa H, Van Liempt H;
 DR N-PSDB; Q13608.
 PT DNA encoding amino:adipyl-cysteiny-valine synthetase - used for
 PT prodn. of the enzyme or enhanced prodn. of new or known
 PT beta-lactam antibiotic cpds.
 PS Claim 1; Page 20; 54pp; English.
 CC The DNA sequence was obtd. from five subclones isolated from a
 CC gene library of A. chrysogenum C10 (ATCC 48). The protein
 CC sequence was deduced from the DNA. Three distinct regions of
 CC homology have been identified, domains I, II and III. Within
 CC these domains several even more conserved elements can be
 CC distinguished. Since the enzyme synthesises a tripeptide, which
 CC most probably requires the activation of three amino acids, a
 CC role for these domains in the activation reactions seems likely.
 CC A fourth domain is thought to act as a thioesterase.
 CC The gene can be used to express the synthetase enzyme which can
 CC be used for the prodn. of new beta-lactam antibiotics.
 CC See also R13896.
 SQ Sequence 3712 AA;
 SQ 262A; 214R; 144N; 219D; 0 B; 42 C; 180Q; 237E; 0 Z; 237G; 127H;
 SQ 130I; 413L; 135K; 58 M; 131F; 184P; 282S; 217T; 39 W; 128Y; 273V;
 Initial Score = 7 Optimized Score = 7 Significance = 3.37
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X
 DXVXXX
 |
 GRNARLYKTGDLVRWIRHANGDGEIEY
 1740 1750 1760
 14. US-08-121-713B-22 (1-7)
 P90290 Human Duchenne muscular dystrophy gene.

ID P90290 standard; protein; 3685 AA.
 AC P90290;
 DT 10-JAN-1990 (first entry)
 DE Human Duchenne muscular dystrophy gene.
 KW Duchenne muscular dystrophy; protein deletion; antiserum.
 OS Homo sapiens.
 PN EP-331514-A.
 PD 06-SEP-1989.
 PF 03-MAR-1989; 302145.
 PR 04-MAR-1988; JP-051313.
 PA (AJIN) Ajinomoto Co., Inc.
 PI Ishiguro T, Eguchi C.
 DR WPI; 89-257828/36.
 PT Detecting human Duchenne muscular dystrophy - by assaying for associated
 PT protein deletion or defect using antibody to the protein.
 PS Claim 3; page 15-16; 20pp; English.
 CC Peptides may be prepared consisting of all or part of the sequence and
 CC used to produce Abs for detecting protein deletions or defects in the
 CC gene.
 SQ Sequence 3685 AA;
 SQ 213 A; 197 R; 167 N; 193 D; 0 B; 36 C; 306 Q; 388 E; 0 Z; 112 G; 88 H;
 SQ 170 I; 457 L; 294 K; 78 M; 80 F; 132 P; 243 S; 199 T; 70 W; 69 Y; 191 V;
 SQ 2 Others;
 Initial Score = 7 Optimized Score = 7 Significance = 3.37
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X
 DXVXXX
 |
 ALRVLQNNVLDLVNIGSTDIVDGNHKL
 80 90 X 100
 15. US-08-121-713B-22 (1-7)
 P90373 Sequence encoded by human muscular dystrophy (MD)
 ID P90373 standard; Protein; 3685 AA.
 AC P90373;
 DT 29-MAR-1992 (first entry)
 DE Sequence encoded by human muscular dystrophy (MD) cDNA.
 KW Dystrophin; muscular dystrophy; probe; antibody; diagnosis;
 KW prenatal; heterozygote; gene therapy; genetic screening;
 KW foetal screening.
 OS Homo sapiens.
 PN WO8906286-A.
 PD 13-JUL-1989.
 PF 16-DEC-1988; U04504.
 PR 22-DEC-1987; US-136618.
 PA (CHIL-) CHILDRENS MED CENT.
 PI Kunkel LM, Monaco A, Hoffman EP, Koenig M;
 DR N-PSDB; N90338.
 PT Muscular dystrophy gene - used for prepn. of probes, dystrophin
 PT polypeptide and antibodies for diagnosis and therapy of muscular
 PT dystrophy
 PS Disclosure; Fig 8; 68pp; English.
 CC The inventors claim an MD probe comprising a purified ss NA SQ which

1

4923 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Score	Sig.	Frame
1. S18135	Calcium release channel - Pig	5034	7	7	2.78	0
2. S27272	ryanodine receptor, brain - r	4872	7	7	2.78	0
3. S18252	perlecan - mouse	3707	7	7	2.78	0
4. S28916	dystrophin - mouse	3678	7	7	2.78	0
5. S43048	polyketide synthase - Strepto	3519	7	7	2.78	0
6. A46105	polyprotein(NS1, NS3, NS5, NS	3415	7	7	2.78	0
7. S44887	ZK112.7 protein - Caenorhabdi	3343	7	7	2.78	0
8. S37536	macrogolgin - human	3259	7	7	2.78	0
9. S20473	fatty-acid synthase (EC 2.3.1	3104	7	7	2.78	0
10. S28645	Fatty-acid synthase (EC 2.3.1	3104	7	7	2.78	0
11. S42373	hypothetical protein - Caenor	3051	7	7	2.78	0
12. S33642	homeotic protein zfh-2 - frui	3005	7	7	2.78	0
13. S17796	inositol-trisphosphate recept	2701	7	7	2.78	0
14. S11661	talin - mouse	2541	7	7	2.78	0
15. S30446	fatty-acid synthase (EC 2.3.1	2505	7	7	2.78	0
16. S46955	protein-tyrosine-phosphatase	2466	7	7	2.78	0
17. S42612	transmembrane protein precurs	2437	7	7	2.78	0
18. S44861	DNA topoisomerase II - Caenor	2434	7	7	2.78	0
19. S48905	Highly similar to C. elegans	2413	7	7	2.78	0
20. A46299	disabled product (alternative	2411	7	7	2.78	0
21. S46578	hypothetical protein YBR1012	2368	7	7	2.78	0
22. S41121	acetyl-CoA carboxylase (EC 6.	2339	7	7	2.78	0
23. S44625	C50C3.6 protein - Caenorhabdi	2329	7	7	2.78	0
24. S45306	notch 3 protein - mouse	2318	7	7	2.78	0
25. S16166	hexabrachion - human	2203	7	7	2.78	0
26. S14015	tenascin - human	2199	7	7	2.78	0
27. S33124	tpr protein - human	2094	7	7	2.78	0
28. S26058	probable transforming protein	2090	7	7	2.78	0
29. S39796	aggrekan - chicken	2071	7	7	2.78	0
30. S12050	protein-tyrosine-phosphatase	1997	7	7	2.78	0
31. S27356	aggrekan - chicken	1951	7	7	2.78	0
32. B43963	RNA viral polymerase - Hantav	1951	7	7	2.78	0
33. S47771	hypothetical protein - yeast	1928	7	7	2.78	0
34. S37771	ankyrin, erythrocyte - mouse	1848	7	7	2.78	0
35. S14568	microtubule-associated protei	1828	7	7	2.78	0
36. S13507	microtubule-associated protei	1825	7	7	2.78	0
37. S34131	microtubule-associated protei	1824	7	7	2.78	0
38. S44849	K12H4.8 protein - Caenorhabdi	1822	7	7	2.78	0
39. S41729	polyketide synthase - Strepto	1803	7	7	2.78	0
40. S44920	ZK688.5 protein - Caenorhabdi	1799	7	7	2.78	0
41. S13178	6-methylsalicylate decarboxyl	1774	7	7	2.78	0
42. S24600	projectin - fruit fly (Drosop	1742	7	7	2.78	0
43. A37491	orf1 putative helicase/polyme	1737	7	7	2.78	0
44. S18644	multifunctional amino acid--t	1714	7	7	2.78	0
45. S42369	Clathrin heavy chain homolog	1681	7	7	2.78	0

1. US-08-121-713B-22 (1-7)

S18135 Calcium release channel - Pig (fragment)

ENTRY S18135 #type fragment
 TITLE Calcium release channel - Pig (fragment)
 ORGANISM #formal name Sus scrofa domestica #common name domestic pig
 DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change 22-Nov-1993
 ACCESSIONS S18135
 REFERENCE S18135 Harbitz, I.; Kristensen, T.; Bosnes, M.; Kran, S.; Davies, W.

#submission submitted to the EMBL Data Library, October 1991
 #accession S18135
 #status preliminary
 #residues 1-5034 #label HAR
 #cross-references EMBL:X62880
 SUMMARY #length 5034 #checksum 8168
 SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 2.78
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X X
 DXVXXXX
 |
 |
 AILRSIVPLDILVGIISLPQIPTLGG
 2460 X 2470 2480

2. US-08-121-713B-22 (1-7)

S27272 ryanodine receptor, brain - rabbit

ENTRY S27272 #type complete
 TITLE ryanodine receptor, brain - rabbit
 ORGANISM #formal name Oryctolagus cuniculus #common name domestic rabbit
 DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change 22-Nov-1993

ACCESSIONS S27272
 REFERENCE S27272 Hakamata, Y.; Nakai, J.; Takeshima, H.; Imoto, K.
 #authors FEBS Lett. (1992) 312:229-235
 #journal Primary structure and distribution of a novel ryanodine receptor/calcium release channel from rabbit brain.

#accession S27272 preliminary
 #status 1-4872 #label HAK
 #residues 1-4872
 #cross-references EMBL:X68650
 SUMMARY #length 4872 #molecular-weight 551925 #checksum 4706
 SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 2.78
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X X
 DXVXXXX
 |
 |

SILRSIVPTEDLVGIISLPKLPSLNNK
 2330 2340

3. US-08-121-713B-22 (1-7)

S18252 perlecan - mouse

ENTRY S18252 #type complete
 TITLE perlecan - mouse
 ORGANISM #formal name Mus musculus #common name house mouse
 DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change 22-Nov-1993

ACCESSIONS S18252
REFERENCE S18252
#authors Noonan, D.M.; Fulle, A.; Valente, P.; Cai, S.; Horigan, E.;
Sasaki, M.; Yamada, Y.; Hassell, J.R.
#journal J. Biol. Chem. (1991) 266:22939-22947
#title The complete sequence of perlecan, a basement membrane
heparan sulfate proteoglycan, reveals extensive similarity
with laminin A chain, low density lipoprotein-receptor, and
the neural cell adhesion molecule.
#cross-references MUID:92078153
#accession S18252 preliminary
#status 1-3707 #label NOO
##residues
##cross-references EMBL:M77174
SUMMARY #length 3707 #molecular-weight 398291 #checksum 1636
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 2.78
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXX
| |
QRGSIQVDGEDLVTRGPGPNVAVNTK
3620 3630 X 3640

4. US-08-121-713B-22 (1-7)
S28916 dystrophin - mouse
ENTRY S28916 #type complete
TITLE dystrophin - mouse
ORGANISM #formal name Mus musculus #common name house mouse
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993

ACCESSIONS S28916
REFERENCE S28916
#authors Bies, R.D.; Phelps, S.F.; Cortez, M.D.; Roberts, R.; Caskey,
C.F.; Chamberlain, J.S.
#journal Nucleic Acids Res. (1992) 20:1725-1731
#title Human and murine dystrophin mRNA transcripts are
differentially expressed during skeletal muscle, heart, and
brain development.

#accession S28916 preliminary
#status 1-3678 #label BIE
##residues
##cross-references EMBL:M68859
SUMMARY #length 3678 #molecular-weight 425815 #checksum 2460
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 2.78
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXX
| |
ALRVLQNNVDLVNIGSTDIVDGNHKL
80 90 X 100

5. US-08-121-713B-22 (1-7)
S43048 polyketide synthase - Streptomyces antibioticus

ENTRY S43048 #type complete
TITLE polyketide synthase - Streptomyces antibioticus
ORGANISM #formal name Streptomyces antibioticus
DATE 20-Oct-1994; #sequence_revision 20-Oct-1994; #text_change
20-Oct-1994

ACCESSIONS S43048
REFERENCE S43048
#authors Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Salas,
J.A.

#submission submitted to the EMBL Data Library, February 1993
#accession S43048 preliminary
#status 1-3519 #label SWA
##residues
##cross-references EMBL:L09654
SUMMARY #length 3519 #molecular-weight 368532 #checksum 1912
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 2.78
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXX
| |
RRGGSAPGAGDLVRELEGLGGARVSVR
1240 X 1250 1260

6. US-08-121-713B-22 (1-7)
A46105 polyprotein(NS1, NS3, NS5, NS2A, NS2B, NS4A, NS4B,

ENTRY A46105 #type complete
TITLE polyprotein(NS1, NS3, NS5, NS2A, NS2B, NS4A, NS4B, C-small
capsid protein, E=large envelope protein,
(pr)M-membrane-anchored protein precursor) - Powassan virus
ORGANISM #formal name Powassan virus
DATE 07-Apr-1994; #sequence_revision 07-Apr-1994; #text_change
07-Apr-1994

ACCESSIONS A46105
REFERENCE A46105
#authors Mandl, C.W.; Holzmann, H.; Kunz, C.; Heinz, F.X.
#journal Virology (1993) 194:173-184
#title Complete genomic sequence of Powassan virus: evaluation of
genetic elements in tick-borne versus mosquito-borne
flaviviruses.

#cross-references MUID:93242744
#accession A46105 preliminary
#status A46105
##molecule type genomic RNA
##residues 1-3415 #label MAN
##cross-references NCBI:130654; NCBI:130655
#note sequence extracted from NCBI backbone
SUMMARY #length 3415 #molecular-weight 378568 #checksum 9967
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 2.78
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXXX
| |

RDIPYLPKTDLVCSLVGRKERAWEA
3350 3360 X 3370

7. US-08-121-713B-22 (1-7)
S44887 ZK112.7 protein - Caenorhabditis elegans

ENTRY S44887 #type complete
TITLE ZK112.7 protein - Caenorhabditis elegans
ORGANISM #formal name Caenorhabditis elegans
DATE 14-Sep-1994; #sequence_revision 14-Sep-1994; #text_change

ACCESSIONS S44887
REFERENCE S44613
#authors Du, Z.
#submission submitted to the EMBL Data Library, May 1993
#accession S44887

##status preliminary
##residues 1-3343 ##label DUA
##cross-references EMBL:L14324
SUMMARY #length 3343 #molecular-weight 375748 #checksum 6289
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 2.78
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXXX
| |

PTQSELINHVLDLVSTQNSDKPFMTL
1070 X 1080 1090

8. US-08-121-713B-22 (1-7)
S37536 macrogolgin - human

ENTRY S37536 #type complete
TITLE macrogolgin - human
ORGANISM #formal name Homo sapiens #common name man
DATE 09-Dec-1993; #sequence_revision 09-Dec-1993; #text_change

ACCESSIONS S37536
REFERENCE S37536
#authors Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.;
Griffiths, G.; Renz, M.

#submission submitted to the EMBL Data Library, September 1993
#accession S37536
##status preliminary
##residues 1-3259 ##label SEE
##cross-references EMBL:X75304
SUMMARY #length 3259 #molecular-weight 376075 #checksum 4495
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 2.78
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXXX
| |

AEERVAELARDLVEMEOKLLMVTKENK
2710 2720

9. US-08-121-713B-22 (1-7)
S20473 fatty-acid synthase (EC 2.3.1.85) - Brevibacterium

ENTRY S20473 #type complete
TITLE fatty-acid synthase (EC 2.3.1.85) - Brevibacterium
ORGANISM #formal name Brevibacterium ammoniagenes
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change

ACCESSIONS S20473
REFERENCE S20473
#authors Meurer, G.; Biermann, G.; Schuetz, A.; Harth, S.; Schweizer, E.

#journal Mol. Gen. Genet. (1992) 232:106-116
#title Molecular structure of the multifunctional fatty acid synthetase gene of Brevibacterium ammoniagenes: its sequence of catalytic domains is formally consistent with a head-to-tail fusion of the two yeast genes FAS1 and FAS2.

#cross-references MUID:92204122
#accession S20473
##status preliminary
##residues 1-3104 ##label MEU
##cross-references EMBL:X64795
SUMMARY #length 3104 #molecular-weight 327466 #checksum 3966
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 2.78
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXXX
| |

SEADDNAAVVDLVTSELGADWPLVSP
1970 X 1980

10. US-08-121-713B-22 (1-7)
S28645 Fatty-acid synthase (EC 2.3.1.85) - Brevibacterium

ENTRY S28645 #type complete
TITLE Fatty-acid synthase (EC 2.3.1.85) - Brevibacterium
ORGANISM #formal name Brevibacterium ammoniagenes
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change

ACCESSIONS S28645
REFERENCE S28645

```
#authors      Schweizer, E.
#submission  submitted to the EMBL Data Library, December 1992
#accession   S28645
#status      preliminary
#residues    1-3104 #label SCH
##cross-references EMBL:N64795
#length 3104 #molecular-weight 327452 #checksum 4026
SUMMARY
SEQUENCE
```


The scores below are sorted by initial score.
Significance is calculated based on initial score.

4077 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig.	Frame
1. YHP9 YEAST	HYPOTHETICAL 433.2 KD PROTEIN	3744	7	7	2.35	0
2. YOG7 CAEEL	HYPOTHETICAL 375.7 KD PROTEIN	3343	7	7	2.35	0
3. YX3 CAEEL	HYPOTHETICAL 337.6 KD PROTEIN	3051	7	7	2.35	0
4. ZFH2 DROME	ZINC-FINGER PROTEIN 2 (ZINC-F	3005	7	7	2.35	0
5. YLJ6 CAEEL	HYPOTHETICAL 272.0 KD PROTEIN	2329	7	7	2.35	0
6. YCF2 EPIVI	HYPOTHETICAL 260 KD PROTEIN (2216	7	7	2.35	0
7. YEO2 YEAST	HYPOTHETICAL 264.2 KD PROTEIN	2163	7	7	2.35	0
8. YCF2 SPIOL	HYPOTHETICAL 250 KD PROTEIN (2131	7	7	2.35	0
9. YHD0 YEAST	HYPOTHETICAL 210.4 KD PROTEIN	1868	7	7	2.35	0
10. YH68 CAEEL	HYPOTHETICAL 208.3 KD PROTEIN	1822	7	7	2.35	0
11. Y025 CAEEL	HYPOTHETICAL 202.6 KD PROTEIN	1799	7	7	2.35	0
12. YK84 YEAST	HYPOTHETICAL 203.3 KD PROTEIN	1764	7	7	2.35	0
13. YL17 YEAST	HYPOTHETICAL 197.5 KD PROTEIN	1758	7	7	2.35	0
14. YCFX TOBAC	HYPOTHETICAL 199 KD PROTEIN (1708	7	7	2.35	0
15. YK11 YEAST	HYPOTHETICAL 195.2 KD PROTEIN	1683	7	7	2.35	0
16. YHD5 YEAST	PROBABLE ATP-DEPENDENT PERMEA	1592	7	7	2.35	0
17. YCF1 YEAST	METAL RESISTANCE PROTEIN YCF1	1515	7	7	2.35	0
18. YK80 YEAST	HYPOTHETICAL 167.8 KD PROTEIN	1483	7	7	2.35	0
19. YS3 YEAST	HYPOTHETICAL 164.4 KD PROTEIN	1430	7	7	2.35	0
20. YIP9 YEAST	HYPOTHETICAL 156.9 KD PROTEIN	1375	7	7	2.35	0
21. Y0L4 CAEEL	HYPOTHETICAL 152.4 KD PROTEIN	1351	7	7	2.35	0
22. YK5 CAEEL	HYPOTHETICAL 153.4 KD PROTEIN	1330	7	7	2.35	0
23. YX2 XENLA	TRANSPOSON TX1 HYPOTHETICAL 1	1308	7	7	2.35	0
24. YK28 CAEEL	HYPOTHETICAL 143.2 KD PROTEIN	1257	7	7	2.35	0
25. YNCA CAEEL	HYPOTHETICAL 142.5 KD PROTEIN	1254	7	7	2.35	0
26. YED1 YEAST	PROBABLE E1-E2 ATPASE YEL031W	1215	7	7	2.35	0
27. YK76 YEAST	HYPOTHETICAL 137.5 KD PROTEIN	1195	7	7	2.35	0
28. YK08 YEAST	PUTATIVE 128.2 KD TRANSCRIPTI	1170	7	7	2.35	0
29. YK62 YEAST	HYPOTHETICAL 133.3 KD PROTEIN	1157	7	7	2.35	0
30. YL8 YEAST	PROBABLE E1-E2 ATPASE YIL048W	1151	7	7	2.35	0
31. YHUL ECOLI	HYPOTHETICAL 125.7 KD PROTEIN	1139	7	7	2.35	0
32. Y000 YEAST	HYPOTHETICAL 122.6 KD PROTEIN	1113	7	7	2.35	0
33. YLD5 CAEEL	HYPOTHETICAL 127.9 KD PROTEIN	1112	7	7	2.35	0
34. YEP ECOLI	HYPOTHETICAL 123.8 KD PROTEIN	1107	7	7	2.35	0
35. YK5 CAEEL	PUTATIVE SERINE/THREONINE-PRO	1087	7	7	2.35	0
36. YJ84 YEAST	HYPOTHETICAL 121.4 KD PROTEIN	1058	7	7	2.35	0
37. YMG2 MCGE	HYPOTHETICAL 114.4 KD PROTEIN	1052	7	7	2.35	0
38. YK3 CAEEL	PUTATIVE SERINE/THREONINE-PRO	1044	7	7	2.35	0
39. Y043 CAEEL	HYPOTHETICAL 115.4 KD PROTEIN	1040	7	7	2.35	0
40. YHIV ECOLI	HYPOTHETICAL 111.5 KD PROTEIN	1037	7	7	2.35	0
41. YK7 YEAST	HYPOTHETICAL 120.4 KD PROTEIN	1037	7	7	2.35	0
42. YM8 YEAST	HYPOTHETICAL 117.9 KD PROTEIN	1032	7	7	2.35	0
43. YK26 CAEEL	HYPOTHETICAL 113.7 KD PROTEIN	1018	7	7	2.35	0
44. PK2 SCHPO	PROTEIN KINASE C-LIKE 2 (EC 2	1016	7	7	2.35	0
45. DPOL_EBV	DNA POLYMERASE (EC 2.7.7.7).	1015	7	7	2.35	0

- US-08-121-713B-22 (1-7)
YHP9 YEAST HYPOTHETICAL 433.2 KD PROTEIN IN HXT5-CDC12 INTERG

ID YHP9 YEAST STANDARD; PRT; 3744 AA.
AC P38911;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 433.2 KD PROTEIN IN HXT5-CDC12 INTERGENIC REGION.
GN YHR099W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RM 94378003
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,
RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
RA NHAN M., RIFKIN L., RILES L., ST.PETER H., TREVASKIS E., VAUGHAN K.,
RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
RA VAUDIN M.;
RL SCIENCE 265:2077-2082(1994).
DR EMBL; U00060; SCH9332.
DR PIR; S46715; S46715.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 3744 AA; 433171 MW; 20943107 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X X
DXVXXX
|
PLVFALQQYDPDIVSQGLRLELCIDNL
890 X 900

- US-08-121-713B-22 (1-7)
YOG7 CAEEL HYPOTHETICAL 375.7 KD PROTEIN ZK112.7 PRECURSOR IN

ID YOG7 CAEEL STANDARD; PRT; 3343 AA.
AC P34616;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 375.7 KD PROTEIN ZK112.7 PRECURSOR IN CHROMOSOME III.
GN ZK112.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RM 94130718
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SIMS M., SWALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RC NATURE 368:32-38 (1994).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY OF CELL ADHESION
CC MOLECULES. STRONG, TO D.MELANOGASTER FAT TUMOR SUPPRESSOR.
DR EMBL; L14324; CEZK112.
DR PIR; S44887; S44887.
DR WORMPEP; ZK112.7; CE00378.
DR PROSITE; PS00232; CADHERIN.
KW HYPOTHETICAL PROTEIN; CELL ADHESION; SIGNAL; TRANSMEMBRANE;
KW CYTOSKELETON; GLYCOPROTEIN; CALCIUM-BINDING; REPEAT.
FT SIGNAL 1 ?
FT CHAIN ? 3343
FT DOMAIN ? 3228
FT TRANSMEM 3250
FT DOMAIN 3251 3343
FT CARBOHYD 22 22
FT CARBOHYD 149 149
FT CARBOHYD 250 250
FT CARBOHYD 288 288
FT CARBOHYD 369 369
FT CARBOHYD 467 467
FT CARBOHYD 612 612
FT CARBOHYD 752 752
FT CARBOHYD 806 806
FT CARBOHYD 941 941
FT CARBOHYD 966 966
FT CARBOHYD 970 970
FT CARBOHYD 985 985
FT CARBOHYD 1042 1042
FT CARBOHYD 1335 1335
FT CARBOHYD 1425 1425
FT CARBOHYD 1429 1429
FT CARBOHYD 1557 1557
FT CARBOHYD 1563 1563
FT CARBOHYD 1597 1597
FT CARBOHYD 1624 1624
FT CARBOHYD 1695 1695
FT CARBOHYD 1702 1702
FT CARBOHYD 1895 1895
FT CARBOHYD 1900 1900
FT CARBOHYD 2053 2053
FT CARBOHYD 2129 2129
FT CARBOHYD 2203 2203
FT CARBOHYD 2382 2382
FT CARBOHYD 2391 2391
FT CARBOHYD 2410 2410
FT CARBOHYD 2414 2414
FT CARBOHYD 2431 2431
FT CARBOHYD 2527 2527
FT CARBOHYD 2530 2530
FT CARBOHYD 2564 2564
FT CARBOHYD 2621 2621
FT CARBOHYD 2665 2665
FT CARBOHYD 2712 2712
FT CARBOHYD 2798 2798
FT CARBOHYD 2809 2809

FT CARBOHYD 2927 2927 POTENTIAL.
FT CARBOHYD 2976 2976 POTENTIAL.
FT CARBOHYD 3045 3045 POTENTIAL.
FT CARBOHYD 3222 3222 POTENTIAL.
FT CARBOHYD 3225 3225 POTENTIAL.
SQ SEQUENCE 3343 AA; 375745 MW; 22105723 CN;
Initial Score = 7 Optimized Score = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X X
DXVXXX
PTQSELIHVDIVSTQNSDMRPFMMTL
1070 X 1080 1090
3. US-08-121-713B-22 (1-7)
YNX3 CAEEL HYPOTHETICAL 337.6 KD PROTEIN T20G5.3 IN CHROMOSOM
ID YNX3 CAEEL STANDARD; PRT; 3051 AA.
AC P34576;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 337.6 KD PROTEIN T20G5.3 IN CHROMOSOME III.
GN T20G5.3.
OS CAENORHABDITIS ELEGANS.
OC EURARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA BERKS M., SMITH A.;
RL SUBMITTED (DEC-1993) TO THE SWISS-PROT DATA BANK.
CC -!- SIMILARITY: TO EGF AND CMP REPEATS.
DR EMBL; Z30423; CET20G5.
DR PIR; S42373; S42373.
DR WORMPEP; T20G5.3; CE00478.
DR PROSITE; PS00022; EGF.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 3051 AA; 337581 MW; 20021622 CN;
Initial Score = 7 Optimized Score = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X X
DXVXXX
VQTCTPKQKTDIVFLIDGSGSIGSYVF
510 X 520 530
4. US-08-121-713B-22 (1-7)
ZFH2_DROME ZINC-FINGER PROTEIN 2 (ZINC-FINGER HOMEODOMAIN PRO
ID ZFH2 DROME STANDARD; PRT; 3005 AA.
AC P28167;
DT 01-OCT-1994 (REL. 30, CREATED)

DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE ZINC-FINGER PROTEIN 2 (ZINC-FINGER HOMEODOMAIN PROTEIN 2).
GN ZFH-2.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RM 92001539
RA FORTINI M.E., LAI Z., RUBIN G.M.;
RL MECH. DEV. 34:113-122(1991).
CC -1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
CC NERVOUS SYSTEM.
CC -1- TISSUE SPECIFICITY: LARGELY RESTRICTED TO THE CNS OF LATE EMBRYO.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: CONTAINS THREE HOMEODOMAIN DOMAINS.
DR EMBL; M63450; DMZFH2.
DR PIR; S27817; S27817.
DR HSP; P02836; S33642.
DR FLYBASE; FBGN0004607; ZFH2.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING; HOMEODOMAIN; NUCLEAR PROTEIN;
KW REPEAT.
FT ZN FING 133 156 C2H2-TYPE.
FT ZN FING 559 582 C2H2-TYPE.
FT ZN FING 614 638 C2H2-TYPE.
FT ZN FING 732 756 C2H2-TYPE.
FT ZN FING 897 916 C2H2-TYPE (DEGENERATE).
FT ZN FING 940 964 C2H2-TYPE.
FT ZN FING 999 1023 C2H2-TYPE.
FT ZN FING 1074 1098 C2H2-TYPE.
FT ZN FING 1210 1233 C2H2-TYPE.
FT ZN FING 1341 1365 C2H2-TYPE.
FT ZN FING 1438 1462 C2H2-TYPE.
FT ZN FING 1477 1500 C2H2-TYPE (DEGENERATE).
FT ZN FING 1513 1535 C2H2-TYPE.
FT ZN FING 1541 1564 C2H2-TYPE.
FT DNA BIND 1797 1856 HOMEODOMAIN 1.
FT DNA BIND 2154 2213 HOMEODOMAIN 2.
FT ZN FING 2234 2256 C2H2-TYPE.
FT ZN FING 2371 2393 C2H2-TYPE.
FT DNA BIND 2760 2819 HOMEODOMAIN 3.
SQ SEQUENCE 3005 AA; 332056 MW; 22430417 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X X
DXVXXX
|
SSSFLKQOQLVDPPEQCLTNQWADT
2010 2020 2030

5. US-08-121-713B-22 (1-7)
YLJ6_CAEEL HYPOTHETICAL 272.0 KD PROTEIN C50C3.6 IN CHROMOSOM

ID YLJ6_CAEEL STANDARD; PRT; 2329 AA.
AC P34369;

DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 272.0 KD PROTEIN C50C3.6 IN CHROMOSOME III.
GN C50C3.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RM STRAIN=BRISTOL N2;
RM 94150718
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKIN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SMALLDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA WATSON R., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RL NATURE 368:32-38(1994).
DR EMBL; L14433; CEC50C3.
DR PIR; S44625; S44625.
DR WORMPEP; C50C3.6; CE00122.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 2329 AA; 272025 MW; 19536537 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X X
DXVXXX
|
LYRLANQLLTLVDNDNYFYLFDMKSF
230 240 X 250

6. US-08-121-713B-22 (1-7)
YCF2 EPVI HYPOTHETICAL 260 KD PROTEIN (ORF 2216).

ID YCF2 EPVI STANDARD; PRT; 2216 AA.
AC P30072;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 260 KD PROTEIN (ORF 2216).
GN YCF2.
OS EPIFAGUS VIRGINIANA (BEECHDROPS).
OG CHLOROPLAST; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC EUKARYOTA; METAZOA; SCOPHULARIALES; OROBANCHACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RM 93066301
RA WOLFE K.H., MORDEN C.W., PALMER J.D.;
RL PROC. NATL. ACAD. SCI. U.S.A. 89:10648-10652(1992).
CC -1- SIMILARITY: TO A SIMILAR ORF IN OTHER PLANTS CHLOROPLAST.

DR EMBL; M81884; CHEVCG.
KW CHLOROPLAST; HYPOTHETICAL PROTEIN.
SQ SEQUENCE 2216 AA; 259511 MW; 21025832 CN;
Initial Score = 7 Optimized Score = 7 Significance = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXXX
|
DLFTLSITEPLVHKGFSPSYMDQK
630 640

7. US-08-121-713B-22 (1-7)
YE02_YEAST HYPOTHETICAL 264.2 KD PROTEIN IN RAD3-BMH1 INTERGE

ID YE02_YEAST STANDARD; PRT; 2163 AA.
AC P32639;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE HYPOTHETICAL 264.2 KD PROTEIN IN RAD3-BMH1 INTERGENIC REGION.
GN YER172C OR SYGP-ORF66.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGEN R., OEFNER P., OH C.,
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;
RA SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 1-169 FROM N.A.
RA MULLIGAN J.T., DIETRICH F.S., HENNESSEY K.M., SEHL P., KOMP C.,
RA WEI Y., TAYLOR P., NAKAHARA K., ROBERTS D., DAVIS R.W.;
RA SUBMITTED (FEB-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO HUMAN HYPOTHETICAL MYELOBLAST PROTEIN (D29641).
DR EMBL; U18922; U18922.
DR EMBL; L11229; SC5YGP5.
DR PIR; S30856; S30856.
KW HYPOTHETICAL PROTEIN; ATP-BINDING.
FT NP_BIND 79 86 ATP (POTENTIAL).
FT NP_BIND 521 528 ATP (POTENTIAL).
SQ SEQUENCE 2163 AA; 246183 MW; 18812673 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXXX
|
TEKWKVAKGLNDLVEQYKFRFETHSKRE

380 X 390 400
8. US-08-121-713B-22 (1-7)
YCF2_SPIOL HYPOTHETICAL 250 KD PROTEIN (ORF 2131).

ID YCF2_SPIOL STANDARD; PRT; 2131 AA.
AC P08973;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 250 KD PROTEIN (ORF 2131).
GN YCF2.
OS SPINACIA OLERACEA (SPINACH).
OG CHLOROPLAST.
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC CAROPHYLLALES; CHENOPODIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RM 88295221
RA ZHOU D.X., MASSENET O., QUIGLEY F., MARION M.J., MONEGER F.,
RA HUBER P., MACHE R.;
RL CURR. GENET. 13:433-439(1988).
CC -1- SIMILARITY: TO A SIMILAR ORF IN OTHER PLANTS CHLOROPLAST.
DR EMBL; X07908; CHSO2131.
DR PIR; S01446; S01446.
KW CHLOROPLAST; HYPOTHETICAL PROTEIN.
SQ SEQUENCE 2131 AA; 249964 MW; 19992229 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXXX
|
GSITMGSNVRDLVAFINEALSISITQK
1680 X 1690

9. US-08-121-713B-22 (1-7)
YHD0_YEAST HYPOTHETICAL 210.4 KD PROTEIN IN GUT1-RIM1 INTERGE

ID YHD0_YEAST STANDARD; PRT; 1868 AA.
AC P38737;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 210.4 KD PROTEIN IN GUT1-RIM1 INTERGENIC REGION.
GN YH1030W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RM 94378003
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,
RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,

RA NHAN M., RIFKIN L., RILES L., ST. PETER H., TREVASKIS E., VAUGHAN K.,
 RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
 RA VAUDIN M.;
 RL SCIENCE 265:2077-2082 (1994).
 DR EMBL; U11583; SCH9196.
 DR PIR; S48938; S48938.
 KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
 FT TRANSEM 695 715 POTENTIAL.
 FT TRANSEM 747 767 POTENTIAL.
 SQ SEQUENCE 1868 AA; 210430 MW; 17651645 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.35
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X
 DXVXXXX
 |
 VRNFALTLLIDLVKHSFGAIKPTPKL
 1300 X 1310 1320

10. US-08-121-713B-22 (1-7)
 YM68_CAEEL HYPOTHETICAL 208.3 KD PROTEIN K12H4.8 IN CHROMOSOM

ID YM68 CAEEL STANDARD; PRT; 1822 AA.
 AC P34529;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 208.3 KD PROTEIN K12H4.8 IN CHROMOSOME III.
 GN K12H4.8.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RM 94150718
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
 RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
 RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
 RA WATERSTON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WOHLDMAN P.;
 RL NATURE 368:32-38 (1994).
 CC -1- SIMILARITY: TO EUKARYOTIC INITIATION FACTOR-4A AND RIBONUCLEASE
 III.
 DR EMBL; L14331; CEK12H4.
 DR PIR; S44849; S44849.
 DR WORMPEP; K12H4.8; CE00273.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 1822 AA; 208291 MW; 15724531 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.35
 Residue Identity = 28% Matches = 2 Mismatches = 5

Gaps = 0 Conservative Substitutions = 0

X X
 DXVXXXX
 |
 AMDSVIETASDLVLSKYGAKPYEVVI
 210 X 220

11. US-08-121-713B-22 (1-7)
 YO25_CAEEL HYPOTHETICAL 202.6 KD PROTEIN ZK688.5 IN CHROMOSOM

ID YO25 CAEEL STANDARD; PRT; 1799 AA.
 AC P34675;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 202.6 KD PROTEIN ZK688.5 IN CHROMOSOME III.
 GN ZK688.5.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RM 94150718
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
 RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
 RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
 RA WATERSTON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WOHLDMAN P.;
 RL NATURE 368:32-38 (1994).
 DR EMBL; L16621; CEZK688.
 DR PIR; S44920; S44920.
 DR WORMPEP; ZK688.5; CE00463.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 1799 AA; 202641 MW; 15440747 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.35
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X
 DXVXXXX
 |
 DKKTRDVIESDLVKMRGYKRNIGSRNQ
 1610 X 1620

12. US-08-121-713B-22 (1-7)
 YKB4_YEAST HYPOTHETICAL 203.3 KD PROTEIN IN PUT3-CC1 INTERGE

ID YKB4 YEAST STANDARD; PRT; 1764 AA.
 AC P34241; P34242;
 DT 01-FEB-1994 (REL. 28, CREATED)

DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 203.3 KD PROTEIN IN PUT3-CCE1 INTERGENIC REGION.
GN YKLO14C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RM 94205264
RA WIEMANN S., VOSS H., SCHWAGER C., RUPP T., STEGEMANN J.,
RA ZIMMERMANN J., GROTHUES D., SENSEN C., ERLE H., HEWITT N.,
RA BANREVI A., ANSORGE W.;
RL YEAST 9:1343-1348(1993).
RN [2]
RP REVISIONS.
RA WIEMANN S., VOSS H., SCHWAGER C., RUPP T., STEGEMANN J.,
RA ZIMMERMANN J., GROTHUES D., SENSEN C., ERLE H., HEWITT N.,
RA BANREVI A., ANSORGE W.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X74152; SCPUKGA.
DR EMBL; Z28014; SCYKLO14C.
DR PIR; S37827; S37827.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1764 AA; 203286 MW; 16178809 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXX
| |
TNNYDATTTCDLVKYDDTSGVDM
1400 X 1410

13. US-08-121-713B-22 (1-7)
Y1R7_YEAST HYPOTHETICAL 197.5 KD PROTEIN IN SUC2 5'REGION.

ID Y1R7_YEAST STANDARD; PRT; 1758 AA.

AC P40434;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 197.5 KD PROTEIN IN SUC2 5'REGION.
GN Y1L177C.

OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;

RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA CHURCHER C.M., CONNOR R., COPESEY T., DEAR S., DEVLIN K., FRASER A.,
RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA WALSH S.V., WHITEHEAD S.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: TO HYPOTHETICAL PROTEIN IN SUBTELOMERIC Y' REPEAT
(246921).

DR EMBL; Z47047; SCCHRIX.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1758 AA; 197511 MW; 15509879 CN;
Initial Score = 7 Optimized Score = 7 Significance = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXX
| |
EFESALNNKNDIVHCPSITLIFEISIPTE
20 X 30 X 40

14. US-08-121-713B-22 (1-7)
YCFX_TOBAC HYPOTHETICAL 199 KD PROTEIN (ORF 1708).

ID YCFX_TOBAC STANDARD; PRT; 1708 AA.

AC P09977;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 199 KD PROTEIN (ORF 1708).
GN YCF2.

OS NICOTIANA TABACUM (COMMON TOBACCO).

OG CHLOROPLAST.

OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC SOLANALES; SOLANACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. BRIGHT YELLOW 4;

RA SUGIURA M.;

RL SUBMITTED (AUG-1986) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP COMPLETE GENOME.

RA SHINOZAKI K., OHME M., TANAKA M., WAKASUGI T., HAYASHIDA N.,

RA MATSUBAYASHI T., ZAITA N., CHUNWONGSE J., OBOKATA J.,

RA YAMAGUCHI-SHINOZAKI K., OHTO C., TORAZAWA K., MENG B.Y., SUGITA M.,

RA DENO H., KAMOGASHIRA T., YAMADA K., KUSUDA J., TAKAIWA F., KATO A.,

RA TOHDOH N., SHIMADA H., SUGIURA M.;

RL EMBO J. 5:2043-2049(1986).

CC -!- FUNCTION: NOT YET KNOWN.

CC -!- SIMILARITY: TO A SIMILAR ORF IN OTHER PLANTS CHLOROPLAST.

CC -!- CAUTION: IN TOBACCO THIS ORF IS IN TWO PARTS: ORF 581 (N-TERMINAL)

CC AND ORF 1708 (C-TERMINAL). IT COULD BE DUE TO A FRAMESHIFT ERROR.

DR EMBL; Z00044; CHNTXX.

DR PIR; A05205; A05205.

KW CHLOROPLAST; HYPOTHETICAL PROTEIN.

SQ SEQUENCE 1708 AA; 199282 MW; 15156803 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXX
| |
DLFTLSITPDVLVYHKGFAPSIDSCGL
50 X 60 X 70

15. US-08-121-713B-22 (1-7)
YK11_YEAST HYPOTHETICAL 195.2 KD PROTEIN IN GCN3-DAL80 INTERG

ID YK11_YEAST STANDARD; PRT; 1683 AA.
AC P36126;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 195.2 KD PROTEIN IN GCN3-DAL80 INTERGENIC REGION.
GN YKR031C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA URESTARAZD L.A., JAUNIAUX J.-C.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Z28256; SCYKR031C.
DR PIR; S38103; S38103.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1683 AA; 195203 MW; 14046134 CN;

Initial Score = 7 Optimized Score = 2.35
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
DXVXXX

LMREHLGCDVLDLVEFVEKKFERFEKFA
1160 1170 X 1180

maryh@stic

stdin

NeWSprinter20

Fri May 19 11:10:08 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5

24

Use these
all search
"X" amino
acids

100000
N
U50000
M
B
E
P

Scores:	Mean	Median	Standard Deviation
	0	1	1.09
Times:	CPU	Total Elapsed	
	00:00:30.06	00:00:31.00	

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.

The scores below are sorted by initial score. Significance is calculated based on initial score.

63 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
1. R11749	Human alpha-2 macroglobulin b	1484	8	8	7.32 0
2. R11334	Recombinant human alpha-2 mac	1474	8	8	7.32 0
3. R32356	Excitatory amino acid receptor	956	8	8	7.32 0
4. R41232	GAT-3 transporter.	632	8	8	7.32 0
5. R41229	GAT-3 GABA transporter.	627	8	8	7.32 0
6. R49590	GAT-B transporter.	627	8	8	7.32 0
7. R51267	Sequence of human prostagland	604	8	8	7.32 0
8. R51268	Sequence of murine prostaglan	604	8	8	7.32 0
9. R36533	Gp85-37 clone 18 prod.	585	8	8	7.32 0
10. R40213	Sequence of a 90K tumour-asso	585	8	8	7.32 0

11. R40172 Sequence of an immunoregulin
12. R41359 Tumour associated 90K antigen
13. R40784 Sequence of cyclophilin assoc
14. R26843 GDP dissociation stimulatory
15. R07466 Polypeptide with enzymatic ac
16. R13984 Phytoene dehydrogenase-4H
17. R13985 Phytoene dehydrogenase-4H fro
18. P93630 Sequence of rat transin.
19. R11704 Cytolysis Inhibitor.
20. P60326 Interleukin-1 gene product.
21. P60680 Prepro- and mature interleuki
22. R14855 Mature Interleukin-1.
23. P81197 Sequence encoded by human per
24. R15747 Interleukin-1 beta with 116 a
25. R15748 Interleukin-1 beta with 116 a
26. P50043 Sequence of interleukin-1 (IL
27. R42213 Human Interleukin-1.
28. R06358 Monkey IL-2 beta deduced from
29. P70306 Sequence of interleukin 1-bet
30. P60307 Sequence of new antitumour su
31. R42447 Human interleukin-1 beta prec
32. P61521 Sequence of new antitumour su
33. P61520 Sequence of new antitumour su
34. P81873 Partial sequence of rabbit st
35. R10613 Hybrid IL-1 beta/alpha-1 anti
36. P94906 Sequence of amino-terminal ex
37. R03065 Human interleukin-1-beta expr
38. P70525 Interleukin-2 encoded by plas
39. R23668 Interleukin-1 beta Ser 63, Se
40. R23669 Interleukin-1 beta Ala 126 Al
41. P90473 Human interleukin-1-beta modi
42. R04209 Interleukin - 1 beta sequence
43. R05386 Polypeptide with physiological
44. R06269 Radioactive iodine labeled in
45. R07286 Interleukin-1-beta deriv.

1. US-08-121-713B-24 (1-8)
R11749 Human alpha-2 macroglobulin bait region mutant.

ID R11749 standard; Protein; 1484 AA.
AC R11749;
DT 04-JUN-1991 (first entry)
DE Human alpha-2 macroglobulin bait region mutant.
KW Human alpha-2 macroglobulin; growth media; proteinase inhibitor.
OS Homo sapiens.
FH Region 701..759
FT /label= P2P bait region
FN Key Location/Qualifiers
PD WO9103557-A.
PE 21-MAR-1991.
PF 29-AUG-1990; DK0225.
PR 29-AUG-1989; DK-004237.
PR 29-AUG-1989; DK-004236.
PR 29-AUG-1989; DK-004236.
PA (NOVO) NOVO NORDISK A/S.
PI Esper B, Lars S-J;
DR WPI; 91-102075/14.
PT DNA encoding alpha macro-globulin - used to produce recombinant

PT protein for use in growth media, proteinase inhibitors and as
PT carriers
PS Disclosure; page 47; 78pp; English.
CC This alpha-2 macroglobulin bait mutant comprises a bait region from
CC the human pregnancy zone protein (P2P). It has a proteinase inhib-
CC itor profile similar to that of the human P2P. It is useful as an
CC additive to growth-media and as a carrier in gene- and enzyme rep-
CC lacement-therapy as well as being a proteinase inhibitor..
CC See also Q11176.
SQ Sequence 1484 AA;
SQ 95 A; 44 R; 66 N; 51 D; 0 B; 25 C; 75 Q; 101E; 0 Z; 94 G; 37 H;
SQ 58 I; 144L; 89 K; 24 M; 61 F; 82 P; 125S; 101T; 12 W; 58 Y; 142V;
Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X
XXXXXXX
PKGNRLAOWQSFQLEGGLKQFSPLSSE
170 180 190
2. US-08-121-713B-24 (1-8)
R11334 Recombinant human alpha-2 macroglobulin.
ID R11334 standard; Protein; 1474 AA.
AC R11334;
DT 04-JUN-1991 (first entry)
DE Recombinant human alpha-2 macroglobulin.
KW Human alpha-2 macroglobulin; growth media; proteinase inhibitor.
OS Homo sapiens.
PN WO9103557-A.
PD 21-MAR-1991.
PF 29-AUG-1990; DK0225.
PR 29-AUG-1989; DK-004237.
PR 29-AUG-1989; DK-004235.
PR 29-AUG-1989; DK-004236.
PA (NOVO) NOVO NORDISK A/S.
PI Esper B, Lars S-J;
DR WPI; 91-102075/14.
PT DNA encoding alpha macro-globulin - used to produce recombinant
PT protein for use in growth media, proteinase inhibitors and as
PT carriers
PS Disclosure; page 34; 78pp; English.
CC This recombinant human alpha-2 macroglobulin is useful as an add-
CC itive to growth-media, as a proteinase inhibitor and as a carrier
CC in gene- and enzyme replacement-therapy. See also Q11581.
SQ Sequence 1474 AA;
SQ 94 A; 46 R; 63 N; 53 D; 0 B; 25 C; 73 Q; 104E; 0 Z; 91 G; 41 H;
SQ 57 I; 143L; 89 K; 26 M; 62 F; 76 P; 123S; 101T; 11 W; 56 Y; 140V;
Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X
XXXXXXX

PKGNRIAQWQSFOLEGLKQFSFPLSSE
170 180 190

3. US-08-121-713B-24 (1-8)
R32356 Excitatory amino acid receptor 1.

ID R32356 standard; Protein; 956 AA.
AC R32356;
DT 21-JUN-1993 (first entry)
DE Excitatory amino acid receptor 1.
KW EAAR-1; brain; kainate; glutamate; assay; CNS.
FT Key Location/Qualifiers
FT Peptide 1..20
FT /note= "signal peptide"
FT Protein 21..956
FT /note= "mature protein"
FN EP-529994-A.
PD 03-MAR-1993.
PF 25-AUG-1992; 307723.
PR 26-AUG-1991; US-750090.
PA (ALIE-) ALLELIX BIOPHARMACEUTICALS INC.
PI Kamboj R, Nutt SL, Shekter L, Wosnick MA;
DR WPI; 93-069001/09.
DR N-PSDB; Q36915.
PT Polynucleotide sequence encoding human excitatory amino acid-1
PT receptor - for assaying cpds. which bind to the receptor in
PT screening for new CNS drugs
PS Disclosure; Fig 1; 28pp; English.
CC Human hippocampal cDNA was subjected to PCR amplification using
CC primers corresp. to regions of the rat GluR1 gene. The amplified
CC prod. was labelled and used to probe a cDNA library from the same
CC source and 50 positive clones isolated; one contained a 1058 bp
CC sequence. This was labelled and used to screen a different
CC hippocampal cDNA library to identify two overlapping clones from
CC which the 67 kb phagemid pBS/HumEPIa (ATCC 75063) contg. the entire
CC open reading frame for EAAR1 receptor was constructed. This
CC complete insert could be removed as a 3.7 kb NotI fragment. The
CC sequence encodes the excitatory amino acid-1 receptor (R) or a
CC kainate binding fragment of it. (R) binds glutamate and has ligand
CC binding properties characteristic of Kainate-type EAA receptors. Cells
CC contg. the gene may be used to assay cpds. for possible use in treatment
CC of CNS disorders.
SQ Sequence 956 AA;
SQ 71 A; 59 R; 42 N; 34 D; 0 B; 24 C; 39 Q; 65 E; 0 Z; 61 G; 19 H;
SQ 58 I; 105L; 33 K; 28 M; 45 F; 45 P; 71 S; 54 T; 15 W; 28 Y; 60 V;

Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X X
XXXXXXX

ACLICAKAECLINLEKILRQFLISKDTIL
170 X 180

4. US-08-121-713B-24 (1-8)

R41232 GAT-3 transporter.

ID R41232 standard; Protein; 632 AA.
AC R41232;
DT 22-MAR-1994 (first entry)
DE GAT-3 transporter.
KW GABA transporter; gamma-aminobutyric acid; taurine transporter;
KW translation inhibition; monoclonal antibodies; transgenic animals;
KW cell membranes; epilepsy; anxiety; migraine; ischaemia.
OS Homo sapiens.
PN WO9318143-A.
PD 16-SEP-1993.
PF 04-MAR-1993; U01959.
PR 04-MAR-1992; US-847742.
PR 13-OCT-1992; US-959936.
PA (SYNA-) SYNAPTIC PHARM CORP.
PI Borden LA, Hartig PR, Smith KE, Weinshank RL;
DR WPI; 93-303437/38.
DR N-PSDB; Q48764.
PT New mammalian transporters for GABA or taurine - are used in
PT drugs for treating epilepsy, anxiety ischaemia and form
PT antibodies for detecting presence of the transporters on cell
PT surface
PS Claim 1; Fig 10B; 218pp; English.
CC The sequences (Q48760-61) encode novel mammalian GABA transporters,
CC sequences (Q48763-64) encode human GABA transporters. The sequences
CC can be used as probes to detect specific mRNA (i.e expression of
CC the transporter) and also for diagnosing a predisposition to disease
CC associated with expression of a specific allele. Sequence (Q48762)
CC shows a related taurine transporter gene.
SQ Sequence 632 AA;
SQ 48 A; 20 R; 24 N; 15 D; 0 B; 27 C; 10 Q; 33 E; 0 Z; 55 G; 9 H;
SQ 49 I; 67 L; 26 K; 20 M; 38 F; 29 P; 35 S; 35 T; 19 W; 32 Y; 41 V;

Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X X
XXXXXXX

VVFICCGIPVFFLEALGQFTSEGGIT
100 X 110 120

5. US-08-121-713B-24 (1-8)
R41229 GAT-3 GABA transporter.

ID R41229 standard; Protein; 627 AA.
AC R41229;
DT 22-MAR-1994 (first entry)
DE GAT-3 GABA transporter.
KW GABA transporter; gamma-aminobutyric acid; taurine transporter;
KW translation inhibition; monoclonal antibodies; transgenic animals;
KW cell membranes; epilepsy; anxiety; migraine; ischaemia.
OS Rattus rattus.
PN WO9318143-A.
PD 16-SEP-1993.
PF 04-MAR-1993; U01959.
PR 04-MAR-1992; US-847742.

PR 13-OCT-1992; US-959936.
 PA (SYNA-) SYNAPTIC PHARM CORP.
 PI Borden LA, Hartig PR, Smith KE, Weinshank RL;
 DR WPI; 93-303457/38.
 DR N-PSDB; Q48761.
 PT New mammalian transporters for GABA or taurine - are used in
 PT drugs for treating epilepsy, anxiety ischaemia and form
 PT antibodies for detecting presence of the transporters on cell
 PT surface
 PS Claim 1; Fig 1B; 218pp; English.
 CC The sequences (Q48760-61) encode novel mammalian GABA transporters.
 CC sequences (Q48763-64) encode human GABA transporters. The sequences
 CC can be used as probes to detect specific mRNA (i.e. expression of
 CC the transporter) and also for diagnosing a predisposition to disease
 CC associated with expression of a specific allele. Sequence (Q48762)
 CC shows a related taurine transporter gene.
 SQ Sequence 627 AA;
 SQ 47 A; 22 R; 23 N; 14 D; 0 B; 25 C; 11 Q; 34 E; 0 Z; 58 G; 7 H;
 SQ 45 I; 71 L; 26 K; 18 M; 40 F; 27 P; 33 S; 33 T; 19 W; 32 Y; 42 V;

Initial Score = 8 Optimized Score = 8 Significance = 7.32
 Residue Identity = 0% Matches = 0 Mismatches = 8
 Gaps = 0 Conservative Substitutions = 0

X X
 XXXXXXXX

VVFICGIPVFFLETALGQFTSEGGIT
 90 100 X 110

6. US-08-121-713B-24 (1-8)

R49590 GAT-B transporter.

ID R49590 standard; Protein; 627 AA.
 AC R49590;
 DT 15-AUG-1994 (first entry)
 DE GAT-B transporter.
 KW GAT-B transporter; GABA transporter; neuron.
 OS Rattus sp.

FH Key Location/Qualifiers
 FT Region 54..74
 FT /note= "putative transmembrane region"
 FT Region 81..101
 FT /note= "putative transmembrane region"
 FT Region 126..146
 FT /note= "putative transmembrane region"
 FT Misc difference 148
 FT /note= "potential N-glycosylation site"
 FT Misc difference 151
 FT /note= "potential N-glycosylation site"
 FT Misc difference 159
 FT /note= "potential N-glycosylation site"
 FT Region 223..239
 FT /note= "putative transmembrane region"
 FT Region 248..270
 FT /note= "putative transmembrane region"
 FT Region 297..317
 FT /note= "putative transmembrane region"
 FT Region 330..351

FT /note= "putative transmembrane region"
 FT Region 383..403
 FT /note= "putative transmembrane region"
 FT Region 433..450
 FT /note= "putative transmembrane region"
 FT Region 468..488
 FT /note= "putative transmembrane region"
 FT Misc difference 505
 FT /note= "potential phosphorylation site for Ca2+
 FT calmodulin-dependent protein-kinase II"
 FT Region 509..528
 FT /note= "putative transmembrane region"
 FT Region 548..566
 FT /note= "putative transmembrane region"
 FT Misc difference 598
 FT /note= "potential phosphorylation site for
 FT protein-kinase C"
 FT Misc difference 603
 FT /note= "potential phosphorylation site for Ca2+
 FT calmodulin-dependent protein-kinase II"
 FT WO9404573-A.
 PN 03-MAR-1994.
 PD 12-AUG-1993; U07576.
 PF 14-AUG-1992; US-930078.
 PR (UYOR-) UNIV OREGON HEALTH SCI.
 PI Amara SG, Clark JA;
 DR WPI; 94-083116/10.
 DR P-PSDB; Q44335.
 PT DNA encoding a GAT-B transporter expressed in neurons but not in
 PT glia - used to confer GAT-B expression on non-GAT-B producing
 PT cells and render cells capable of GABA uptake
 PS Disclosure; Fig 1; 44pp; English.
 CC A sense primer (given in Q44336) and antisense primer (Q44337) were
 CC used for PCR amplification of GAT-B cDNA isolated from a rat mid-
 CC brain genomic library. A cDNA clone was sequenced (Q44335) and
 CC encoded GAT-B transporter (R49590). Expression of GAT-B mRNA in rat
 CC CNS was demonstrated by in situ hybridization using antisense
 CC (Q44338) and sense (Q44339) probes.
 SQ Sequence 627 AA;
 SQ 49 A; 21 R; 23 N; 14 D; 0 B; 25 C; 11 Q; 34 E; 0 Z; 59 G; 7 H;
 SQ 45 I; 68 L; 26 K; 18 M; 40 F; 25 P; 33 S; 33 T; 19 W; 32 Y; 45 V;

Initial Score = 8 Optimized Score = 8 Significance = 7.32
 Residue Identity = 0% Matches = 0 Mismatches = 8
 Gaps = 0 Conservative Substitutions = 0

X X
 XXXXXXXX

VVFICGIPVFFLETALGQFTSEGGIT
 90 100 X 110

7. US-08-121-713B-24 (1-8)

R51267 Sequence of human prostaglandin G/H synthase-2 (PG

ID R51267 standard; Protein; 604 AA.
 AC R51267;
 DT 12-OCT-1994 (first entry)
 DE Sequence of human prostaglandin G/H synthase-2 (PGHS-2).

KW Prostaglandin; hormone; eicosanoid; fatty acid metabolism.
OS Homo sapiens.
PN WO9406919-A.
PD 31-MAR-1994.
PF 22-SEP-1993; U09167.
PR 22-SEP-1992; US-949780.
PR 01-DEC-1992; US-983835.
PR 22-MAR-1993; US-034143.
PR 28-APR-1993; US-054364.
PA (UVRP) UNIV ROCHESTER.
PI Obanion MK, Winn VD, Young DA;
DR WPI; 94-118468/14.
DR N-PSDB; Q61790.
DR New prostaglandin G/H synthase-2 gene - used for producing
PT transgenic cell lines for testing ability of cpds. to inhibit
PT synthesis of prostaglandin(s)
PS Example; Figure 1; 76pp; English.
CC RNA was isolated from a human fibroblast cell line (WI38). PCR
CC primers specific for the human PGHS-1 and PGHS-2 sequences were
CC engineered to amplify the coding regions of either one transcript or
CC the other (see Q61792-95). PCR products of about 2 kb were generated.
CC Three PGHS-2 clones were sequenced in both directions. The clone
CC comprising the PGHS-2 sequence disclosed in Q61790 was selected for
CC transfection. This sequence differs from the human PGHS-2 sequence
CC disclosed by Hla and Heilison, PNAS, 89, 7384 (1992) due to a Glu
CC rather than a Gly at AA posn. 165. Mouse PGHS-2 also has a Glu at
CC this posn.
SQ Sequence 604 AA;
SQ 32 A; 28 R; 29 N; 26 D; 0 B; 14 C; 33 Q; 34 E; 0 Z; 34 G; 19 H;
SQ 35 I; 60 L; 35 K; 14 M; 38 F; 40 P; 34 S; 34 T; 6 W; 26 Y; 33 V;
Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X X
XXXXXXX

YNYQQFIYNNLSILLEGITQVFESFTRQ
390 400 X 410

8. US-08-121-713B-24 (1-8)
R51268 Sequence of murine prostaglandin G/H synthase-2 (P

ID R51268 standard; Protein; 604 AA.
AC R51268;
DT 12-OCT-1994 (first entry)
DE Sequence of murine prostaglandin G/H synthase-2 (PGHS-2).
KW Prostaglandin; hormone; eicosanoid; fatty acid metabolism;
KW griPGHS.
OS Mus musculus.
FH Key Location/Qualifiers
FT Cleavage site 17..18
FT /label= signal peptide cleavage site
FT Modified site 54..56
FT /label= potential N-glycosylation site
FT /note= "see also 130-132,396-398,380-382"
FT Modified site 516
FT /label= putative aspirin modified Ser

PN WO9406919-A.
PD 31-MAR-1994.
PF 22-SEP-1993; U09167.
PR 22-SEP-1992; US-949780.
PR 01-DEC-1992; US-983835.
PR 22-MAR-1993; US-034143.
PR 28-APR-1993; US-054364.
PA (UVRP) UNIV ROCHESTER.
PI Obanion MK, Winn VD, Young DA;
DR WPI; 94-118468/14.
DR N-PSDB; Q61791.
DR New prostaglandin G/H synthase-2 gene - used for producing
PT transgenic cell lines for testing ability of cpds. to inhibit
PT synthesis of prostaglandin(s)
PS Example; Figure 1; 76pp; English.
CC cDNA was prepd. from poly-A enriched RNA from C127 cells of mouse
CC fibroblasts. A directionally cloned cDNA library was constructed in
CC lambda ZAP II. Several positive plaques were isolated and analysed.
CC One about 4.1 kb in length was fully sequenced. This clone encodes
CC a 70 kDa protein. The full sequence has been deposited in GenBank,
CC accession number M88242. Comparison of the 4.1 kb sequence with
CC that previously cloned 2.8 kb PGHS cDNA from mice revealed a
CC single ORF with 64% AA identity to the protein encoded by the 2.8
CC kb PGHS cDNA. The deduced protein sequences are collinear except
CC that the 4.1 kb cDNA has a shorter amino-terminus and longer
CC carboxy-terminus. 'X' in R51268 denotes the translation of codon
CC TAN in Q61790, following the sequence as published.
SQ Sequence 604 AA;
SQ 29 A; 27 R; 30 N; 25 D; 0 B; 13 C; 30 Q; 38 E; 0 Z; 38 G; 18 H;
SQ 34 I; 60 L; 35 K; 14 M; 37 F; 39 P; 34 S; 37 T; 6 W; 26 Y; 33 V;
SQ 1 Others;
Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X X
XXXXXXX

YSFKQFLYNNLSILLEGHGLTQVFESFTRQ
390 400 X 410

9. US-08-121-713B-24 (1-8)
R36533 Gp85-97 clone 18 prod.

ID R36533 standard; Protein; 585 AA.
AC R36533;
DT 19-AUG-1993 (first entry)
DE Gp85-97 clone 18 prod.
KW Glycoprotein; lectin; Mac-2; PHA; lymphocytes; leucoagglutinating;
KW phytohemagglutinin; cancer; amplification; PCR.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..18
FT /note= "leader sequence"
FT Protein 19..585
FT /note= "mature gp85-97"
PN WO9308215-A.
PD 29-APR-1993.

PF 15-OCT-1992; 008878.
PR 16-OCT-1991; US-777121.
PR 15-OCT-1992; US-961404.
PA (CETU) CETUS ONCOLOGY CORP.
PI Casipit CL, Halenbeck R, Kothe KE, Taylor EW, Wang AM;
DR WFI; 93-152424/18.
PR New glyco:protein complex binding to human lectin Mac-2 - also
FT interferes with PHA activation of lymphocytes for treating and
PT preventing cancer, infectious diseases, etc.
PS Claim 1; Page 41; 57pp; English.
CC SK-BR-3 gp97 recovered in partially proteolysed form was denatured
CC and reduced and the 97 and 70 kD mols. were purified using size
CC exclusion HPLC in 0.1 percent SDS. The 97 and 70 kD mols. were
CC digested with Lys-C protease and the resulting peptides purified and
CC sequenced. The N-terminal sequence of gp97 was used to design
CC degenerate primers for use in a PCR reaction on SK-BR-3 mRNA. PCR
CC reactions using primer 2 in combination with primers 1 and 3 gave prods.
CC of 97 and 121 bp. Further PCR using primers 2, 3 and 4 yielded DNA
CC sequences of 740 and 765 respectively. A DNA sequence was obtd. from
CC the PCR prods. Two additional oligonucleotide sequences were
CC synthesised based on this sequence and used to probe a TBP-1 cDNA
CC library to obtain the full-length cDNA sequence that encodes SK-BR-3
CC gp97. Both strands of clone 218 were sequenced. It encodes a novel
CC glycoprotein binds the Mac-2 lectin and interferes with PHA activation
CC of lymphocytes. It can be used for treating or preventing diseases that
CC result from binding of a disease-causing agent to the cell surface of a
CC target cell. The gp. can be used in treatment of cancers, partic.
CC breast cancer. See also R36532.
SQ Sequence 585 AA;
SQ 48 A; 29 R; 17 N; 33 D; 0 B; 16 C; 29 Q; 31 E; 0 Z; 38 G; 9 H;
SQ 14 I; 72 L; 19 K; 9 M; 30 F; 28 P; 54 S; 36 T; 16 W; 22 Y; 35 V;

Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXXX

DLIYAYAVATGDALLEKICLOFLAWNFEA
270 X 280

10. US-08-121-713B-24 (1-8)
R40213 Sequence of a 90K tumour-associated antigen, IR-95

ID R40213 standard; Protein; 585 AA.
AC R40213;
DT 03-FEB-1994 (first entry)
DE Sequence of a 90K tumour-associated antigen, IR-95, capable
DE of binding to SP-2 (Accession No. I-1083).
KW Tumour-associated antigen; breast cancer cell line CG-5;
KW 90 K antigen.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= signal
FT Region 24..125
FT /label= SCRC homology region
FT Binding_site 69

FT /label= potential glycosylation site
FT /note= "see also AAs 125,192,363,398,551,580"
PN WO9316181-A.
PD 19-AUG-1993.
PF 17-FEB-1993; E00385.
PR 17-FEB-1992; IT-RM0099.
PA (UYCH-) UNIV CHIETI ANNUNZIO G D.
PA (UYNV) UNIV NEW YORK STATE.
PI Iacobelli S, Natoli C, Schlessinger J;
DR WPI; 93-272885/34.
DR N-PSDB; Q48170.
PT New 90K tumour-associated antigen, IR-95 - is for use in
PT diagnosis and therapy of cancer, HIV and auto-immune diseases
PS Claim 2; pages 45-48; 68pp; English.
CC 90K antigen has an apparent molecular weight of approx. 95 kd. It is
CC a tumour associated antigen and is elevated in the serum of patients
CC with cancer and also in patients with HIV. It reacts with MAB SP-2
CC which was prod. by immunising mice with proteins that had been
CC released into tissue culture fluid by human MCF-7 breast cancer
CC cells. MAB SP-2 cell line is deposited at the Institut Pasteur,
CC Paris, Accession number 1-1083. 90K is also present in normal
CC subjects. It is purified from the culture fluid of the human breast
CC cancer cell line, CG-5. N-terminal sequencing of the 90K antigen
CC was used to design a 'guesser' nucleotide sequence (Q48171) as a
CC probe to screen a lambda-gt10 library prepd. from MCF7 polyA+ RNA.
CC The complete nucleotide sequence of isolated clones is given in
CC Q48170.

SQ Sequence 585 AA;
SQ 49 A; 29 R; 17 N; 32 D; 0 B; 16 C; 30 Q; 29 E; 0 Z; 41 G; 10 H;
SQ 14 I; 71 L; 19 K; 9 M; 30 F; 31 P; 51 S; 34 T; 16 W; 22 Y; 35 V;
Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXXX

DLIYAYAVATGDALLEKICLOFLAWNFEA
270 X 280

11. US-08-121-713B-24 (1-8)
R40172 Sequence of an immunoregulin (IR)-95 polypeptide.

ID R40172 standard; Protein; 585 AA.
AC R40172;
DT 08-FEB-1994 (first entry)
DE Sequence of an immunoregulin (IR)-95 polypeptide.
KW Tumour associated antigen; breast cancer; cell line CG-5;
KW ovarian cancer; immunoregulin-95; IR-95.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..17
FT /label= signal
FT Region 24..125
FT /label= SCRC homology region
FT Modified_site 71
FT /label= Glycosylation site
FT /note= "see also AAs 125,192,362,398,551,580"

PN W09316180-A.
PD 19-AUG-1993.
PF 17-FEB-1993; E00382.
PR 17-FEB-1992; IT-RM0100.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (UYCH-) UNIV CHIETI ANNUNZIO G D.
PI Azam M, Iacobelli S, Natoli C, Sures I, Ullrich A;
DR WPI; 93-272884/34.
DR N-PSDB; Q46888.
PT Recombinant DNA - encoding tumour associated antigen,
PT immuno-regulin-95
PS Disclosure; Fig 1; 69pp; English.
CC Immunoregulin-95 is a 90K tumour-associated antigen purified from
CC the culture fluid of the human breast cancer cell line, CG-5, the
CC serum of a breast cancer patient, or the ascitic fluid from an
CC ovarian cancer patient. The native antigen, which has a mol. wt. of
CC 95 kD, is present as a high mol. wt. complex. Homology in the region
CC of AAs 35-80 of the 90K antigen is found with type I macrophage
CC scavenger receptor; sea urchin speract receptor; and human
CC lymphocyte glycoprotein T1/Leu-1.
SQ Sequence 585 AA;
SQ 49 A; 29 R; 17 N; 32 D; 0 B; 16 C; 30 Q; 29 E; 0 Z; 41 G; 10 H;
SQ 14 I; 71 L; 19 K; 9 M; 30 F; 31 P; 51 S; 34 T; 16 W; 22 Y; 35 V;
Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X X
XXXXXXX

DIYAYAVATGDALLEKLCLOFLAWNFEA
270 X 280

12. US-08-121-713B-24 (1-8)
R41359 Tumour associated 90K antigen.

ID R41359 standard; Protein; 585 AA.
AC R41359;
DT 03-MAR-1994 (first entry)
DE Tumour associated 90K antigen.
KW Antigen; cancer; inflammation; autoimmune disease; viral infection;
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= Signal peptide.
FT Region 24..125
FT /note= "Region homologous with sea urchin speract
FT receptor"
FT Modified site 69
FT /note= "Potential asparagine linked N-glycosylation
FT site"
FT Modified site 125
FT /note= "Potential asparagine linked N-glycosylation
FT site"
FT Modified site 192
FT /note= "Potential asparagine linked N-glycosylation
FT site"
FT Modified site 362

FT /note= "Potential asparagine linked N-glycosylation
FT site"
FT Modified site 398
FT /note= "Potential asparagine linked N-glycosylation
FT site"
FT Modified site 551
FT /note= "Potential asparagine linked N-glycosylation
FT site"
FT Modified site 580
FT /note= "Potential asparagine linked N-glycosylation
FT site"
PN W09317119-A.
PD 02-SEP-1993.
PF 16-FEB-1993; E00379.
PR 17-FEB-1992; IT-RM0099.
PA (UYCH-) UNIV CHIETI ANNUNZIO G D.
PA (UYNV) UNIV NEW YORK STATE.
PI Iacobelli S, Natoli C, Schlessinger J;
DR WPI; 93-288423/36.
DR N-PSDB; Q46161.
PT 90K Tumour-associated antigen - purified from serum of breast
PT cancer patient or ascitic fluid from ovarian cancer patient
PS Claim 1; Figure 1; 73pp; English.
CC The purified 90K antigen or its antigenic determinant containing
CC fragment can be assayed to diagnose a disorder, such as cancer (especially breast or ovarian cancer) or a viral infection,
CC inflammation, autoimmune disease and/or arthritis in a patient.
CC It may also be used in the preparation of an agent for the
CC treatment of these disorders.
SQ Sequence 585 AA;
SQ 49 A; 29 R; 17 N; 32 D; 0 B; 16 C; 30 Q; 29 E; 0 Z; 41 G; 10 H;
SQ 14 I; 71 L; 19 K; 9 M; 30 F; 31 P; 51 S; 34 T; 16 W; 22 Y; 35 V;
Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXX

DIYAYAVATGDALLEKLCLOFLAWNFEA
270 X 280

13. US-08-121-713B-24 (1-8)
R40784 Sequence of cyclophilin associated membrane protei

ID R40784 standard; Protein; 574 AA.
AC R40784;
DT 04-FEB-1994 (first entry)
DE Sequence of cyclophilin associated membrane protein (CAMP-c).
KW Cyclophilin associated membrane protein; CAMP-c.
OS Acornys californicus.
PN W09316183-A.
PD 19-AUG-1993.
PF 08-FEB-1993; U01123.
PR 07-FEB-1992; US-832862.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Friedman JS, Weissman IL;
DR WPI; 93-272887/34.

DR N-PSDB; Q47694.
PT Cyclophilin C-associated membrane proteins and DNA - used for
PT screening for immunomodulatory agents and for diagnosis and
PT therapy
PS Claim 7; Figure 2; 105pp; English.
CC Q47694 is a murine cDNA sequence which encodes the CAMP-c 77KD
CC polypeptide. The full sequence, including untranslated sequences,
CC is shown in Q47695. The cDNA sequence encodes a polypeptide of 64KD
CC which is glycosylated to give a 77 kD protein. The human homologue
CC of the murine CAMP-c gene is identified and isolated by screening a
CC human genomic clone library with a probe comprising a sequence of
CC about at least 20 contiguous nucleotides (or their complement) of
CC the cDNA sequence of Q47694 or Q48695.
SQ Sequence 574 AA;
SQ 40 A; 26 R; 20 N; 27 D; 0 B; 17 C; 28 Q; 33 E; 0 Z; 37 G; 10 H;
SQ 16 I; 70 L; 13 K; 13 M; 25 F; 26 P; 54 S; 41 T; 12 W; 27 Y; 39 V;
Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXX

DLIYARATGDSMLELGVFLAWNFEF
270 X 280

14. US-08-121-713B-24 (1-8)
R26843 GDP dissociation stimulatory protein.

ID R26843 standard; Protein; 558 AA.
AC R26843;
DT 11-FEB-1993 (first entry)
DE GDP dissociation stimulatory protein.
KW Guanosine 5'-diphosphate; mass production.
OS Not known.
FN J04211700-A.
PD 03-AUG-1992.
PF 08-FEB-1991; 018105.
PR 24-MAY-1990; JP-134480.
PA (MITU) MITSUBISHI KASEI CORP.
DR WPI; 92-305013/37.
DR N-PSDB; Q28232.
PT High mol. wt. mass producible protein - accelerates dissociation
PT of guanosine 5'-di-phosphate from complex of GDP and GDP protein
PS binding type low mol guanosine 5'-tri-phosphate.
PS Claim 1; Fig 1; 17pp; Japanese.
CC The sequence is that of a protein which accelerates dissociation of
CC guanosine 5'-di-phosphate (GDP) from a complex of GDP and GDP protein
CC binding type low molecular guanosine 5'-triphosphate (GTP) binding
CC protein. It may be mass produced using genetic engineering techniques.
SQ Sequence 558 AA;
SQ 54 A; 16 R; 26 N; 28 D; 0 B; 12 C; 32 Q; 46 E; 0 Z; 27 G; 14 H;
SQ 32 I; 85 L; 38 K; 18 M; 12 F; 10 P; 39 S; 19 T; 2 W; 2 Y; 46 V;
Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXX

TEMCLVAFGNLAELSSKEQFASTNIAE
150 160 170

15. US-08-121-713B-24 (1-8)
R07466 Polypeptide with enzymatic activity for the conver

ID R07466 standard; protein; 492 AA.
AC R07466;
DT 28-JAN-1991 (first entry)
DE Polypeptide with enzymatic activity for the conversion of phytoene
DE into lycopene.
KW Carotenoid biosynthesis; vitamin A; cancer; food coloring.
OS Erwinia uredovora.
PN EP-393690-A.
PD 24-OCT-1990.
PF 20-APR-1990; 107493.
PR 21-APR-1989; JP-103078.
PR 05-MAR-1990; JP-053255.
PA (KIRI) KIRIN BEER KK.
PI Misawa N, Kobayashi K, Nakamura K;
DR WPI; 90-322212/43.
DR N-PSDB; Q06296.
PT DNA sequences encoding enzymes for carotenoid biosynthesis - for
PT prodn. of carotenoid cpds. e.g. beta-carotene, lycopene,
PT phytoene, etc.
PS Claim 4; Fig 4; 40pp; English.
CC Gene products are useful for the synthesis of carotenoids, useful as
CC food coloring, vitamin A precursor, and possibly in prevention of
CC cancer.
CC See also Q06293-9.
SQ Sequence 492 AA;
SQ 43 A; 28 R; 15 N; 25 D; 0 B; 4 C; 23 Q; 26 E; 0 Z; 41 G; 20 H;
SQ 18 I; 57 L; 17 K; 8 M; 28 F; 25 P; 26 S; 27 T; 6 W; 21 Y; 34 V;
Initial Score = 8 Optimized Score = 8 Significance = 7.32
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXX

SGKVFYNDQTRLEAQIQQFNPRDVEG
100 X 110 120
> O <
O I O IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq24pir.res made by maryh on Fri 19 May 95 10:48:42-PDT.

Query sequence being compared: US-08-121-713B-24 (1-8)
Number of sequences searched: 75511
Number of scores above cutoff: 4749

30. A43488 genome polypeptide - feline c 1287 8 8 7.00 0
31. S18738 pol protein - simian foamy vi 1161 8 8 7.00 0
32. GNLJLK pol polypeptide - simian foamy 1157 8 8 7.00 0
33. A35098 MHC class III histocompatibil 1132 8 8 7.00 0
34. A47058 Fe-regulated RTX cytotoxin ho 1115 8 8 7.00 0
35. A45761 Ca2+-transporting ATPase (EC 1115 8 8 7.00 0
36. A39299 DNA-directed DNA polymerase (1106 8 8 7.00 0
37. PIXRPC inner layer protein VP1 - por 1082 8 8 7.00 0
38. GNWEC genome polypeptide M - cowpea 1046 8 8 7.00 0
39. A31982 Ca2+-transporting ATPase (EC 1043 8 8 7.00 0
40. PWRBMC Ca2+-transporting ATPase (EC 1042 8 8 7.00 0
41. B40812 Ca2+-transporting ATPase (EC 1042 8 8 7.00 0
42. A31981 Ca2+-transporting ATPase (EC 1042 8 8 7.00 0
43. S04652 Ca2+-transporting ATPase (EC 1042 8 8 7.00 0
44. A33881 hyaluronidase - Clostridium p 1042 8 8 7.00 0
45. S27540

1. US-08-121-713B-24 (1-8)

A45086 HC-toxin synthetase - fungus (Cochliobolus carbonu

ENTRY A45086 #type complete
TITLE HC-toxin synthetase - fungus (Cochliobolus carbonum)
ORGANISM #formal name Cochliobolus carbonum
DATE 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 04-Dec-1994
ACCESSIONS A45086
REFERENCE Scott-Craig, J.S.; Panaccione, D.G.; Pocard, J.A.; Walton, J.D.
#journal J. Biol. Chem. (1992) 267:26044-26049
#title The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous fungus Cochliobolus carbonum is encoded by a 15.7-kilobase open reading frame.
#cross-references MUID:93100328
#contents SB11
#accession A45086
##status preliminary
##molecule_type nucleic acid
##residues 1-5232 #label SCO
##cross-references NCBI:120884
##note sequence extracted from NCBI backbone
##note sequence not compared to nucleotide translation
SUMMARY #length 5232 #molecular-weight 576294 #checksum 2412
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXRVHSDPCIEVQLLERLMEQFGHNLQTL
2920 X 2940

2. US-08-121-713B-24 (1-8)

A38905 dynein heavy chain, cytosolic - rat

ENTRY A38905 #type complete
TITLE dynein heavy chain, cytosolic - rat
CONTAINS dynein ATPase (EC 3.6.1.33)
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 15-Apr-1994 #sequence_revision 02-May-1994 #text_change 08-Dec-1994
ACCESSIONS A38905
REFERENCE Zhang, Z.; Tanaka, Y.; Nonaka, S.; Aizawa, H.; Kawasaki, H.; Nakata, T.; Hirokawa, N.
#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:7928-7932
#title The primary structure of rat brain (cytoplasmic) dynein heavy chain, a cytoplasmic motor enzyme.
#accession A38905
##molecule_type mRNA
##residues 1-4644 #label ZHA
##cross-references GB:D13896
##note the nucleotide sequence is not given in this paper
CLASSIFICATION #superfamily dynein heavy chain, cytosolic
KEYWORDS ATP; blocked amino end; heterotetramer; hydrolase; microtubule binding

FEATURE
1904-1911
2222-2229
2593-2600
2935-2942
1910
2228
2599
2941
SUMMARY
SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXEKQVELYRNGQRLLEKQRFPPSWLYI
1190 X 1200

3. US-08-121-713B-24 (1-8)

S38128 dynein heavy chain - yeast (Saccharomyces cerevisi

ENTRY S38128 #type complete
TITLE dynein heavy chain - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YKR054c
ORGANISM #formal name Saccharomyces cerevisiae
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Dec-1994

ACCESSIONS S38128; S43077; S38130; S37701
REFERENCE S38118

#authors Viissers, S.; Drrestarazu, L.A.; Jauniaux, J.C.
#submission submitted to the Protein Sequence Database, March 1994
#accession S38128
##molecule_type DNA
##residues 1-4092 #label VIS

```
##cross-references EMBL:Z28279
REFERENCE S43077
#authors Eshel, D.
#submission submitted to the EMBL Data Library, March 1993
#accession S43077
##molecule_type DNA
##residues 1-2117, IV, 2120-4092 ##label ESH
##cross-references EMBL:Z21877
REFERENCE S38130
#authors van Vliet-Reedijk, J.C.; Planta, R.J.
#submission submitted to the Protein Sequence Database, March 1994
#accession S38130
##molecule_type DNA
##residues 1-787 ##label VAN
##cross-references EMBL:Z28279
REFERENCE S37701
#authors Li, Y.Y.; Yeh, E.Y.; Hays, T.; Bloom, K.S.
#submission submitted to the EMBL Data Library, May 1993
#description Disruption of mitotic spindle orientation in a yeast dynein
mutant.
#accession S37701
##molecule_type DNA
##residues 1-588, 'C', 590-600, 'A', 602-1363, 'A', 1365-2631, 'P',
2633-2657, 2659, 'IGW', 2660-2661 ##label LIY
##cross-references EMBL:LI5626
GENETICS
#gene DYN1; DHC1
#map position 11R
KEYWORDS P-loop; purine nucleotide binding
FEATURE
1796-1803 #region purine nucleotide-binding motif A (P-loop)\
2074-2081 #region purine nucleotide-binding motif A (P-loop)\
2418-2425 #region purine nucleotide-binding motif A (P-loop)\
2760-2767 #region purine nucleotide-binding motif A (P-loop)\
1802 #binding site ATP/GTP (Lys) #status predicted\
2080 #binding site ATP/GTP (Lys) #status predicted\
2424 #binding site ATP/GTP (Lys) #status predicted\
2766 #binding site ATP/GTP (Lys) #status predicted\
SUMMARY #length 4092 ##molecular-weight 471343 #checksum 4662
SEQUENCE
Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X
XXXXXXX
DSLWKIKSSLSTFLERQRPFFYFLG
1520 X 1530 1540
4. US-08-121-713B-24 (1-8)
A48126 GCN4 translational activator GCN1 - yeast (Sacchar
ENTRY A48126 #type complete
TITLE GCN4 translational activator GCN1 - yeast (Saccharomyces
cerevisiae)
ORGANISM #formal name Saccharomyces cerevisiae
DATE 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
```

```
18-Nov-1994
A48126
#authors Marton, M.J.; Crouch, D.; Hinnebusch, A.G.
#journal Mol. Cell. Biol. (1993) 13:3541-3556
#title GCN1, a translational activator of GCN4 in Saccharomyces
cerevisiae, is required for phosphorylation of eukaryotic
translation initiation factor 2 by protein kinase GCN2.
##cross-references NUID:93268304
#accession A48126
##status preliminary
##molecule_type nucleic acid
##residues 1-2672 ##label MAR
##cross-references NCBIP:132672; NCBIP:132673
sequence extracted from NCBI backbone
SUMMARY #length 2672 ##molecular-weight 296695 #checksum 2104
SEQUENCE
Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X
XXXXXXX
NDLLNFYKFKAKPLEPILDQFGLIVLSA
1150 1160 X 1170
5. US-08-121-713B-24 (1-8)
S43940 TOR1 protein - yeast (Saccharomyces cerevisiae)
ENTRY S43940 #type complete
TITLE TOR1 protein yeast (Saccharomyces cerevisiae)
ORGANISM #formal name Saccharomyces cerevisiae
DATE 28-Oct-1994; #sequence_revision 28-Oct-1994; #text_change
28-Oct-1994
ACCESSIONS S43940
REFERENCE S43940
#authors Helliwell, S.B.; Wagner, P.; Kunz, J.; Deuter-Reinhard, M.;
Henriquez, R.; Hall, M.N.
#journal Mol. Biol. Cell (1994) 5:105-118
#title TOR1 and TOR2 are structurally and functionally similar but
not identical phosphatidylinositol kinase homologues in
yeast.
#accession S43940
##status preliminary
##residues 1-2470 ##label HEL
##cross-references EMBL:X74857
SUMMARY #length 2470 ##molecular-weight 281354 #checksum 1019
SEQUENCE
Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X
XXXXXXX
RVAVLWHELWYEGLEDASRQFFVEHNIE
```

1960 X 1970 X 1980

```
6. US-08-121-713B-24 (1-8)
  GNNYED genome polyprotein - encephalomyocarditis virus (s

ENTRY      #type complete
TITLE      genome polyprotein - encephalomyocarditis virus (strain
CONTAINS   EMC-D, diabetogenic)
           coat protein VP1; coat protein VP2; coat protein VP3; coat
           protein VP4; core protein P2-A; core protein P2-B; core
           protein P2-C; core protein P3-A; genome-linked protein VPg;
           protease; RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM   #formal name encephalomyocarditis virus, EMCV
#note      host Homo sapiens (man)
DATE       31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
           04-Dec-1994
ACCESSIONS A31473; A60498
REFERENCE  A94395
#authors   Bae, Y.S.; Eun, H.M.; Yoon, J.W.
#journal   Virology (1989) 170:282-287
#title     Genomic differences between the diabetogenic and
           nondiabetogenic variants of encephalomyocarditis virus.
#cross-references MUID:89243189
#accession A31473
##molecule_type genomic RNA
##residues 1-2292 ##label BAE
##note     the authors translated the codon ATG for residue 1079 as
           Asn and GAC for residue 1564 as Val

REFERENCE  A60498
#authors   Bae, Y.S.; Eun, H.M.; Yoon, J.W.
#journal   Diabetes (1989) 38:316-320
#title     Molecular identification of diabetogenic viral gene.
#accession A60498
##molecule_type genomic RNA
##residues 1-1522, 'D', 1524-2292 ##label BA2
CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein
KEYWORDS       coat protein; core protein; genome-linked protein;
           nucleotidyltransferase; polyprotein; proteinase

FEATURE
1-67       #domain leader peptide #status predicted #label LDP\
68-137     #product coat protein VP4 #status predicted #label VP4\
138-393    #product coat protein VP2 #status predicted #label VP2\
394-624    #product coat protein VP3 #status predicted #label VP3\
625-901    #product coat protein VP1 #status predicted #label VP1\
902-1058   #product core protein P2-A #status predicted #label P2A\
1059-1194  #product core protein P2-B #status predicted #label P2B\
1195-1519  #product core protein P2-C #status predicted #label P2C\
1520-1607  #product core protein P3-A #status predicted #label P3A\
1608-1627  #product genome-linked protein VPg #status predicted
           #label VPg\
1628-1832  #product proteinase #status predicted #label PTS\
1833-2292  #product RNA-directed RNA polymerase #status predicted
           #label RDP
#length 2292 #molecular-weight 255382 #checksum 3920

SUMMARY
SEQUENCE
Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
```

```
X X
XXXXXXX
DAPLPCFONNCLFLEKAGLOFPDNETKE
1470 1480

7. US-08-121-713B-24 (1-8)
  GNNYEB genome polyprotein - encephalomyocarditis virus (s

ENTRY      #type complete
TITLE      genome polyprotein - encephalomyocarditis virus (strain
CONTAINS   EMC-B, nondiabetogenic)
           coat protein VP1; coat protein VP2; coat protein VP3; coat
           protein VP4; core protein P2-A; core protein P2-B; core
           protein P2-C; core protein P3-A; genome-linked protein VPg;
           protease; RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM   #formal name encephalomyocarditis virus, EMCV
#note      host Homo sapiens (man)
DATE       31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
           04-Dec-1994
ACCESSIONS B31473; B60498
REFERENCE  A94395
#authors   Bae, Y.S.; Eun, H.M.; Yoon, J.W.
#journal   Virology (1989) 170:282-287
#title     Genomic differences between the diabetogenic and
           nondiabetogenic variants of encephalomyocarditis virus.
#cross-references MUID:89243189
#accession B31473
##molecule_type genomic RNA
##residues 1-2292 ##label BAE
##note     the authors translated the codon ATG for residue 1079 as
           Asn and GAC for residue 1564 as Val

REFERENCE  A60498
#authors   Bae, Y.S.; Eun, H.M.; Yoon, J.W.
#journal   Diabetes (1989) 38:316-320
#title     Molecular identification of diabetogenic viral gene.
#accession B60498
##molecule_type genomic RNA
##residues 1-15; 17-2292 ##label BA2
CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein
KEYWORDS       coat protein; core protein; genome-linked protein;
           nucleotidyltransferase; polyprotein; proteinase

FEATURE
1-67       #domain leader peptide #status predicted #label LDP\
68-137     #product coat protein VP4 #status predicted #label VP4\
138-393    #product coat protein VP2 #status predicted #label VP2\
394-624    #product coat protein VP3 #status predicted #label VP3\
625-901    #product coat protein VP1 #status predicted #label VP1\
902-1058   #product core protein P2-A #status predicted #label P2A\
1059-1194  #product core protein P2-B #status predicted #label P2B\
1195-1519  #product core protein P2-C #status predicted #label P2C\
1520-1607  #product core protein P3-A #status predicted #label P3A\
1608-1627  #product genome-linked protein VPg #status predicted
           #label VPg\
1628-1832  #product proteinase #status predicted #label PTS\
1833-2292  #product RNA-directed RNA polymerase #status predicted
           #label RDP
#length 2292 #molecular-weight 255495 #checksum 6728

SUMMARY
```

SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXX

DAPLPCFQNNCLFLEKAGLOFRDNRKE
1470 1480

8. US-08-121-713B-24 (1-8)
S35961 capsid polyprotein precursor - encephalomyocarditi

ENTRY S35961 #type complete
TITLE capsid polyprotein precursor - encephalomyocarditis virus
ORGANISM #formal_name encephalomyocarditis virus, EMCV
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993

ACCESSIONS S35961
REFERENCE S35961

#authors Zimmermann, A.; Nelsen-Slax, B.; Kruppenbacher, J.P.; Eggers,
H.J.

#submission submitted to the EMBL Data Library, July 1993

#accession S35961

#status preliminary

#residues 1-2292, ##label ZIM

##cross-references EMBL:X74312

SUMMARY #length 2292 #molecular-weight 255727 #checksum 8853

SEQUENCE

Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXX

EAPLPCFQNNCLFLEKAGLOFRDNRKE
1470 1480

9. US-08-121-713B-24 (1-8)
GNNYE genome polyprotein - encephalomyocarditis virus

ENTRY GNYYE #type complete
TITLE genome polyprotein - encephalomyocarditis virus
CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat
protein VP4; core protein P2-A; core protein P2-B; core
protein P2-C; core protein P3-A; genome-linked protein VPg;
proteinase (EC 3.4.-.-); RNA-directed RNA polymerase (EC
2.7.7.48)

#formal_name encephalomyocarditis virus, EMCV

host Homo sapiens (man)

DATE 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change

08-Dec-1994

ACCESSIONS A03906; JN0383

REFERENCE A03906

#authors

Palmenberg, A.C.; Kirby, E.M.; Janda, M.R.; Drake, N.L.;

Duke, G.M.; Potratz, K.F.; Collett, M.S.

#journal Nucleic Acids Res. (1984) 12:2969-2985

#title The nucleotide and deduced amino acid sequences of the

encephalomyocarditis viral polyprotein coding region.

#cross-references MUID:84169586

#accession A03906

##molecule_type genomic RNA

##residues_ 1-2290 #label PAL

##cross-references GB:X00463

REFERENCE

#authors

Petrov, N.A.; Chizhikov, V.E.; Blinov, V.M.; Karginov, V.A.;

Mikryukov, N.N.; Gutorov, V.V.; Grishaev, M.P.;

Beklemishev, A.B.; Vassilenko, S.K.

#journal Bioorg. Khim. (1984) 10:274-279

#title Nucleotide sequence of the 3'-terminus of

encephalomyocarditis virus RNA.

#cross-references MUID:85022788

#accession JN0383

##molecule_type genomic RNA

##residues_ 1337-1396, 'L', 1398-1517, 'A', 1519-1536, 'E', 1538-1556, 'S',

1558-1611, 'T', 1613-1915, 'N', 1917-1986, 'IH', 1989-2007,

'I', 2009-2048, 'H', 2050-2193, 'K', 2195-2290 #label PET

##cross-references GB:M54935

##note

the authors translated the codon CAU for residue 713 as

Thr and AAC for residue 857 as Asp

CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein

coat protein; core protein; genome-linked protein; hydrolyase;

nucleotidyltransferase; polyprotein; proteinase

FEATURE

1-67

68-136

137-391

392-622

623-910

911-1056

1057-1192

1193-1517

1518-1605

1606-1625

1626-1830

1831-2290

SUMMARY

SEQUENCE

#length 2290 #molecular-weight 255756 #checksum 8698

Initial Score = 8 Optimized Score = 8 Significance = 7.00

Residue Identity = 0% Matches = 0 Mismatches = 8

Gaps = 0 Conservative Substitutions = 0

X X

XXXXXXXX

EAPLPCFQNNCLFLEKAGLOFRDNRKE

1460 1470 1480

10. US-08-121-713B-24 (1-8)

GNVGV genome polyprotein - grapevine fanleaf virus

ENTRY GNVGVG #type complete
TITLE genome polyprotein - grapevine fanleaf virus
CONTAINS 63K protein; 72K protein; genome-linked protein; proteinase
(EC 3.4.-.-); RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM #formal_name grapevine fanleaf virus
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
ACCESSIONS 02-Aug-1994
REFERENCE JQ1373; S15873
#authors Ritzenthaler, C.; Virv, M.; Pinck, M.; Margis, R.; Fuchs, M.;
Pinck, L.
#journal J. Gen. Virol. (1991) 72:2357-2365
#title Complete nucleotide sequence and genetic organization of
grapevine fanleaf nepovirus RNA1.
#cross-references DDBJ:D00915
#accession JQ1373
#molecule_type genomic RNA
#residues 1-2284 #label RIT
#cross-references DDBJ:D00915
REFERENCE S15873
#authors Pinck, M.; Reinbolt, J.; Loudes, A.M.; le Ret, M.; Pinck, L.
#journal FEBS Lett. (1991) 284:117-119
#title Primary structure and location of the genome-linked protein
(VPg) of grapevine fanleaf nepovirus.
#cross-references MUID:91285092
#accession S15873
#molecule_type protein
#residues 1218-1241 #label PIN
GENETICS
#map_position segment 1
CLASSIFICATION #superfamily cowpea aphid-borne mosaic virus genome
polyprotein B
KEYWORDS genome-linked protein; hydrolase; membrane protein;
nucleotidyltransferase; polyprotein; proteinase
FEATURE
1-568 #product 63K protein #status predicted #label STK\
569-1217 #product 72K protein #status predicted #label SRP\
1218-1241 #product genome-linked protein #status experimental
#label GLP\
1242-1460 #product proteinase #status predicted #label TKP\
1461-2284 #product RNA-directed RNA polymerase #status predicted
#label RRP
SUMMARY #length 2284 #molecular-weight 252930 #checksum 8511
SEQUENCE
Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X
XXXXXXX
ISPNVAVKCCVARLEDGIPQFHFWSKYA
1360 1370 X 1380
11. US-08-121-713B-24 (1-8)
S19444 hypothetical protein YCR032W (version 1) - yeast

ENTRY S19444 #type complete
TITLE hypothetical protein YCR032W (version 1) - yeast
(Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein YCR591; hypothetical protein YCR601
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
ACCESSIONS 07-May-1993
REFERENCE S19444; S15052; S19751
S19437
#authors Pohl, F.; Richterich, P.; Wurst, H.
#submission submitted to the Protein Sequence Database, March 1992
#accession S19444
#molecule_type DNA
#residues 1-334 #label POH
#cross-references EMBL:X59720
REFERENCE S15052
#authors Jia, Y.; Slonimski, P.P.; Herbert, C.J.
#journal Yeast (1991) 7:413-424
#title The complete sequence of the unit YCR59, situated between
CRY1 and MAT, reveals two long open reading frames, which
cover 91% of the 10.1 kb segment.
#cross-references MUID:91335897
#accession S15052
#molecule_type DNA
#residues 335-2167 #label JIA
#cross-references EMBL:X59075
REFERENCE S19445
#authors Herbert, C.J.; Jia, Y.; Slonimski, P.P.
#submission submitted to the Protein Sequence Database, March 1992
#accession S19751
#molecule_type DNA
#residues 335-2167 #label HER
#cross-references EMBL:X59720
GENETICS
#map_position 3R
SUMMARY #length 2167 #molecular-weight 250972 #checksum 6407
SEQUENCE
Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X
XXXXXXX
TSNRMTTEKHLYLEIKGQFCISNDNY
260 X 270
12. US-08-121-713B-24 (1-8)
S17478 hypothetical protein YCR032W (version 2) - yeast
ENTRY S17478 #type complete
TITLE hypothetical protein YCR032W (version 2) - yeast
(Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein YCR591; hypothetical protein YCR601
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 07-May-1993 #sequence_revision 07-May-1993 #text_change
ACCESSIONS 02-Aug-1994
S17478; S40917

REFERENCE S17478
#authors Rodriguez, F.; Martegani, E.; Mauri, I.; Alberghina, I.
#journal Yeast (1991) 7:631-641
#title The sequence of 8.8 kb of yeast chromosome III cloned in
lambda PM3270 contains an unusual long ORF (YCR601).
#accession S17478
##molecule_type DNA
##residues 1-2167 ##label ROD
##cross-references EMBL:X62452
REFERENCE S40917
#authors Wicksteed, B.L.; Roberts, A.B.; Saggiocco, F.A.; Brown,
A.J.P.
#journal Yeast (1991) 7:761-772
#title The complete sequence of a 7.5 kb region of chromosome III
from Saccharomyces cerevisiae that lies between CRY1 and
MAT.
#accession S40917
##molecule_type DNA
##residues 1760-1892, 'N', 1894-2167 ##label WIC
##cross-references EMBL:S78624
GENETICS
#map_position 3R #length 2167 #molecular-weight 250870 #checksum 7348
SUMMARY
SEQUENCE
Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X
XXXXXXX

TSNRIMTIERHYLEIKEGQFCISNDNY
260 X 270

13. US-08-121-713B-24 (1-8)
A61231 myosin heavy chain NMHC-A, nonmuscle - human
ENTRY A61231 #type complete
TITLE myosin heavy chain NMHC-A, nonmuscle - human
CONTAINS myosin ATPase (EC 3.6.1.32)
ORGANISM #formal name Homo sapiens #common name man
DATE 12-May-1994 #sequence_revision 14-Jul-1994 #text_change
08-Dec-1994
ACCESSIONS A61231; A34876
REFERENCE A61231
#authors Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa,
K.; Ghula, D.; Adelstein, R.S.; Weir, L.
#journal Circ. Res. (1991) 69:530-539
#title Human nonmuscle myosin heavy chains are encoded by two genes
located on different chromosomes.
#accession A61231
##molecule_type mRNA
##residues 1-715 ##label SIM
##cross-references GB:M69180
REFERENCE A34876
#authors Saez, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:1164-1168
#title Human nonmuscle myosin heavy chain mRNA: generation of

diversity through alternative polyadenylation.
#cross-references MUID:90138958
#accession A34876
##molecule_type mRNA
##residues 715-1961 ##label SAE
##cross-references GB:M31013
GENETICS
#gene GDB:MYH9
#map_position 22q12.3-q13.1
CLASSIFICATION #superfamily myosin heavy chain; myosin head homology
KEYWORDS actin binding; ATP; coiled coil; hydrolase; methylation;
tandem repeat
FEATURE
84-900 #domain myosin head homology #label HEA\
174-181 #region nucleotide-binding motif A (P-loop)\
552-565 #region actin-binding #status predicted\
626-640 #region actin-binding #status predicted\
837-1938 #domain coiled coil #status predicted #label COI\
837-1277 #region S2\
1278-1961 #region light meromyosin\
1939-1961 #domain carboxyl-terminal #label CBT\
125 #modified_site N6,N6,N6-trimethyllysine (Lys) #status
predicted\
180 #binding_site ATP (Lys) #status predicted\
SUMMARY #active_site Cys #status predicted
694,704 #length 1961 #molecular-weight 226741 #checksum 9108
SEQUENCE
Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X
XXXXXXX

ARALEEAEQKAELERLNKQFTMEDL
1490 X 1500 1510

14. US-08-121-713B-24 (1-8)
A33977 myosin heavy chain, cellular - chicken
ENTRY A33977 #type complete
TITLE myosin heavy chain, cellular - chicken
CONTAINS myosin ATPase (EC 3.6.1.32)
ORGANISM #formal name Gallus gallus #common name chicken
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
08-Dec-1994
ACCESSIONS A33977
REFERENCE A33977
#authors Shohet, R.V.; Conti, M.A.; Kawamoto, S.; Preston, Y.A.;
Brill, D.A.; Adelstein, R.S.
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:7726-7730
#title Cloning of the cDNA encoding the myosin heavy chain of a
vertebrate cellular myosin.
#cross-references MUID:90046668
#accession A33977
##molecule_type mRNA
##residues 1-1959 ##label SHO
##cross-references GB:M26510

CLASSIFICATION #superfamily myosin heavy chain; myosin head homology
KEYWORDS actin binding; ATP; coiled coil; hydrolase; methylation; tandem repeat

```

FEATURE
84-836      #domain nucleoside head homology #label HEA\
174-181    #region nucleotide-binding motif A (P-loop)\
552-565    #region actin-binding #status predicted\
626-640    #region actin-binding #status predicted\
837-1936   #domain coiled coil #status predicted #label COI\
837-1277   #region S2\
1278-1959 #region light meromyosin\
1937-1959 #domain carboxyl-terminal #label CBT\
125        #modified site N6,N6,N6-trimethyllysine (Lys) #status
           predicted\
           #binding_site ATP (Lys) #status predicted\
           #active_site Cys #status predicted
length 1959 #molecular-weight 226502 #checksum 3641
SUMMARY
SEQUENCE

```

Initial Score	=	8	Optimized Score	=	8	Significance	=	7.00
Residue Identity	=	0%	Matches	=	0	Mismatches	=	8
Gaps	=	0	Conservative Substitutions	=	0		=	0

[illegible]

ARALEEAEQAELEERVNKQFRTMEDL
1490 X 1500

15. US-08-121-713B-24 (1-8)
A45973 trichohyalin - human

```
ENTRY          A45973          #type complete
TITLE          trichohyalin - human
ORGANISM       #formal_name Homo sapiens #common_name man
DATE           03-May-1994 #sequence_revision 03-May-1994 #text_change
              03-May-1994
ACCESSIONS    A45973
REFERENCE     A45973
AUTHORS       Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry,
              D.A.D.; Steinert, P.M.
JOURNAL       J. Biol. Chem. (1993) 268:12164-12176
#journal      The structure of human trichohyalin. Potential multiple roles
#title        as a functional EF-hand-like calcium-binding protein, a
              cornified cell envelope precursor, and an intermediate
              filament-associated (cross-linking) protein.
```

```

#accession      A45973
#status         preliminary
#molecule_type DNA
#residues       1-1898 #label LEE
#cross-references GB:I09190
#note          authors translated the codon AGG for residue 1714 as Pro
                #length 1898 #molecular-weight 247219 #checksum 4094
SUMMARY

```

```
Initial Score = 8 Optimized Score = 8 Significance = 7.00
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
```

```

X      X
XXXXXXX

EIQOEGKGHGRLLEPGTHQFASVPVRS
1860 X 1870 X 1880

> O <
O | O IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq24spt.res made by on Fri 19 May 95 8:59:52-PDT.
```

Query sequence being compared: US-08-121-713B-24 (1-8)
 Number of sequences searched: 43470
 Number of scores above cutoff: 4385

Results of the initial comparison of US-08-121-713B-24 (1-8) with:
Data bank : Swiss-Prot 31, all entries

Country	Number of Publications
N	10,000
U	5,000
M	1,000
B	100
E	10
R	1
O	1
F	1
S	1
E	1
Q	1
U	1
E	1
N	1
C	1
E	1
S	1

CC -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME, ABLE TO
CC ACTIVATE THE AMINO ACIDS PRO, L-ALA, D-ALA AND AEO (2-AMINO-9,10-
CC EPOXI-8-OXODECANOIC ACID), AND EPIMERIZE PRO AND L-ALA. IT
CC CATALYZES THE PRODUCTION OF HC-TOXIN; A CYCLIC TETRAPEPTIDE.
CC HTS CONSISTS OF HTS-1 AND HTS-2 WHICH ARE PART OF THE SAME
CC POLYPEPTIDE. HTS-1 ACTIVATES L-PROLINE, THIOESTERIFIES L-PRO, AND
CC EPIMERIZES IT TO D-PRO; HTS-2 ACTIVATES AND THIOESTERIFIES L-ALA
CC AND D-ALA AND EPIMERIZES L-ALA TO D-ALA.
CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF HC-TOXIN.
CC -!- COFACTOR: CONTAINS FOUR COVALENTLY BOUND PHOSPHOPANTHETHEINES.
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
DR EMBL; M98024; CCHFS1X.
DR PROSITE; PS00455; AMP BINDING.
KW LIGASE; MULTIFUNCTIONAL ENZYME; PHOSPHOPANTHETHEINE; REPEAT.
FT REPEAT 249 842 DOMAIN 1.
FT REPEAT 1854 2452 DOMAIN 2.
FT REPEAT 3006 3606 DOMAIN 3.
FT REPEAT 4158 4738 DOMAIN 4.
FT BINDING 803 803 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 2414 2414 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 3568 3568 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 4700 4700 PHOSPHOPANTHETHEINE (BY SIMILARITY).
SQ SEQUENCE 5217 AA; 574637 MW; 21594385 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X X
XXXXXXX

RVHSDPCIEVQLLERLMEQFGHNQLIL
2910 X 2920 2930

2. US-08-121-713B-24 (1-8)
DYHC_RAT DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).

ID DYHC RAT STANDARD; PRT; 4644 AA.
AC P38650;
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=BRAIN;
RM 93376715
RA ZHANG Z., TANAKA Y., NONAKA S., AIZAWA H., KAWASAKI H., NAKATA T.,
RA HIROKAWA N.
RL PROC. NATL. ACAD. SCI. U.S.A. 90:7928-7932 (1993).
CC -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN IS
CC THOUGHT TO ACT AS A MOTOR FOR VESICLE TRANSPORT, AND IS THOUGHT
CC TO CONTRIBUTE TO CHROMOSOME MOVEMENTS.
CC -!- SUBUNIT: CONSIST OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF

CC INTERMEDIATE AND LOW MASS POLYPEPTIDES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: TO OTHER DYNEIN HEAVY CHAINS.
DR EMBL; D13896; RNCDDHC.
KW MOTOR PROTEIN; MICROTUBULES; DYNEIN; ATP-BINDING;
KW HEPTAD REPEAT PATTERN.
FT NP BIND 1904 1911 ATP (POTENTIAL).
FT NP BIND 2222 2229 ATP (POTENTIAL).
FT NP BIND 2593 2600 ATP (POTENTIAL).
FT NP BIND 2935 2942 ATP (POTENTIAL).
SQ SEQUENCE 4644 AA; 532240 MW; 23196763 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X X
XXXXXXX

EKQVELYRNGORLLEKORFPPPSWLYI
1190 X 1200

3. US-08-121-713B-24 (1-8)
DYHC_DROME DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).

ID DYHC DROME STANDARD; PRT; 4639 AA.
AC P37276;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).
GN CDHC.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
[1]
RN SEQUENCE FROM N.A.
RA LI M.G., MCGRATH M., SERR M., GEPNER J., HAYS T.;
RL SUBMITTED (AUG-1993) TO EMBL/GENBANK/DDJ DATA BANKS.
CC -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN IS
CC THOUGHT TO ACT AS A MOTOR FOR VESICLE TRANSPORT, AND IS THOUGHT
CC TO CONTRIBUTE TO CHROMOSOME MOVEMENTS.
CC -!- SUBUNIT: CONSIST OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LOW MASS POLYPEPTIDES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: TO OTHER DYNEIN HEAVY CHAINS.
DR EMBL; L23195; DMCYTHA.
DR FLYBASE; FBGN0010349; CDHC.
KW MOTOR PROTEIN; MICROTUBULES; DYNEIN; ATP-BINDING;
KW HEPTAD REPEAT PATTERN.
FT NP BIND 1895 1902 ATP (POTENTIAL).
FT NP BIND 2210 2217 ATP (POTENTIAL).
FT NP BIND 2580 2587 ATP (POTENTIAL).
FT NP BIND 2922 2929 ATP (POTENTIAL).
SQ SEQUENCE 4639 AA; 530152 MW; 22186784 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXDKQVEVFREAQRIILRQRFQFPNTLHV
1180 X 1190 12004. US-08-121-713B-24 (1-8)
DYHC_YEAST DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).

ID DYHC_YEAST STANDARD; PRT; 4092 AA.
AC P36022;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).
GN DYN1 OR DHC1 OR YKRO54C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RM 94068366
RA ESEHL D., URRESTARAZU L.A., VISSERS S., JAUNIAUX J.-C.,
VAN VLIET-REEDIJK J.C., PLANTA R.J., GIBBONS I.R.;
RL PROC. NATL. ACAD. SCI. U.S.A. 90:11172-11176(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA VISSERS S., URRESTARAZU L.A., JAUNIAUX J.-C.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 1-2657 FROM N.A.
RM 94052110
RA LI Y.-Y., YEH E.-Y., HAYS T., BLOOM K.S.;
RL PROC. NATL. ACAD. SCI. U.S.A. 90:10096-10100(1993).
CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN IS
THOUGHT TO ACT IN CYTOPLASMIC MICROTUBULE-BASED MOTILE PROCESSES
INCLUDING VESICLE TRANSPORT, AND CHROMOSOME MOVEMENTS. MAY PLAY AN
IMPORTANT ROLE IN THE PROPER ORIENTATION OF THE MITOTIC SPINDLE
INTO THE BUDDING DAUGHTER CELL YEAST. PROBABLY REQUIRED FOR NORMAL
PROGRESSION OF THE CELL CYCLE.
CC -1- SUBUNIT: CONSIST OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
INTERMEDIATE AND LOW MASS POLYPEPTIDES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. PROBABLY BINDS INDIRECTLY TO
THE INNER PLASMA MEMBRANE.
CC -1- SIMILARITY: TO OTHER DYNEIN HEAVY CHAINS.

DR EMBL; Z21877; SCDFN1A.
DR EMBL; L15626; SCDFC1A.
DR EMBL; Z28279; SCYKRO54C.
DR PIR; S38128; S38128.
DR LISTA; SC00299; DYN1.
KW MOTOR PROTEIN; MICROTUBULES; DYNEIN; ATP-BINDING;
KW HEPTAD REPEAT PATTERN.
FT NP_BIND 1796 1803 ATP (POTENTIAL).
FT NP_BIND 2074 2081 ATP (POTENTIAL).
FT NP_BIND 2418 2425 ATP (POTENTIAL).
FT NP_BIND 2760 2767 ATP (POTENTIAL).
FT NP_BIND 589 589 Y -> C (IN REF. 3).
FT CONFLICT 601 601 V -> A (IN REF. 3).
FT CONFLICT 601 601

FT CONFLICT 1364 1364 E -> A (IN REF. 3).
FT CONFLICT 2118 2119 ML -> IV (IN REF. 1).
FT CONFLICT 2632 2632 A -> P (IN REF. 3).
SQ SEQUENCE 4092 AA; 471337 MW; 21476769 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXDSLKMIKSLSTFLERQRRQRFYFLG
1520 X 1530 15405. US-08-121-713B-24 (1-8)
GCN1_YEAST TRANSLATIONAL ACTIVATOR GCN1.

ID GCN1_YEAST STANDARD; PRT; 2672 AA.
AC P33892;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE TRANSLATIONAL ACTIVATOR GCN1.
GN GCN1.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RM 93268304
RA MARTON M.J., CROUCH D., HINNEBUSCH A.G.;
RL MOL. CELL. BIOL. 13:3541-3556(1993).
CC -1- FUNCTION: TRANSLATIONAL ACTIVATOR OF GCN4. MAY BE INVOLVED IN
SENSING CHARGED tRNA AND STIMULATING THE KINASE ACTIVITY OF GCN2
IN AMINO ACID-STARVED CELLS. REQUIRED IN VIVO FOR THE
PHOSPHORYLATION OF EIF-2-ALPHA ON SERINE-52 BY THE PROTEIN KINASE
GCN2.
CC -1- SIMILARITY: TO ELONGATION FACTOR 3 (EF-3) OF FUNGI.
DR EMBL; L12467; SGCN1A.
DR PIR; A48126; A48126.
DR LISTA; SC00382; GCN1.
KW TRANSLATION REGULATION; ACTIVATOR; REPEAT.
FT DOMAIN 1344 2327 4 X 16 AA APPROXIMATE REPEATS OF R-X-X-
A-X-X-L-X-X-L-V-X-X-X-G.
FT REPEAT 1344 1359 1-1.
FT REPEAT 1624 1639 1-2.
FT REPEAT 1663 1678 1-3.
FT REPEAT 2312 2326 1-4.
FT DOMAIN 1385 1790 2 X 90 AA APPROXIMATE REPEATS.
FT REPEAT 1385 1474 2-1.
FT REPEAT 1701 1790 2-2.
FT DOMAIN 1385 1937 2 APPROXIMATE REPEATS.
FT REPEAT 1385 1674 3-1.
FT REPEAT 1701 1937 3-2.
FT DOMAIN 1624 1937 2 X 51 AA APPROXIMATE REPEATS.
FT REPEAT 1624 1674 4-1.
FT REPEAT 1887 1937 4-2.
SQ SEQUENCE 2672 AA; 296693 MW; 22333332 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXX

NDLNFYKEKAKPLEPILDQGLVLVSA
1150 1160 X 1170

6. US-08-121-713B-24 (1-8)
TOR1_YEAST PHOSPHATIDYLINOSITOL 3-KINASE TOR1 (EC 2.7.1.137)

ID TOR1_YEAST STANDARD; PRT; 2470 AA.
AC P35169;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE PHOSPHATIDYLINOSITOL 3-KINASE TOR1 (EC 2.7.1.137) (PI3-KINASE)
DE (PTDINS-3-KINASE) (PI3K).
GN TOR1 OR DRRL.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RM 94019276
RA CAFFERKEY R., YOUNG P.R., MC LAUGHLIN M.M., BERGSMAN D.J., KOLTIN Y.,
RA SATHE G.M., FACETTE L., ENG W.-K., JOHNSON R.K., LIVI G.P.;
RL MOL. CELL. BIOL. 13:6012-6023(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JK-3D;
RM 94243030
RA HELLWELL S.B., WAGNER P., KUNZ J., DEUTER-REINHARD M., HENRIQUEZ R.,
RA HALL M.N.;
RL MOL. BIOL. CELL 5:105-118(1994).
CC -1- FUNCTION: PHOSPHATIDYLINOSITOL 3-KINASE HOMOLOG REQUIRED FOR G1
CC PROGRESSION. TARGET OF THE ANTIBIOTIC RAPAMYCIN.
CC -1- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL = ADP +
CC 1-PHOSPHATIDYL-ID-MYO-INOSITOL 3-PHOSPHATE.
CC -1- IT MAY ACT ON ANOTHER SUBSTRATE OR PHOSPHORYLATE A DIFFERENT
CC POSITION IN THE PHOSPHATIDYLINOSITOL RING.
CC -1- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
DR EMBL; L19540; SCDRIA.
DR EMBL; X74857; SCTORI.
DR LISTA; SC01314; TOR1.
DR PROSITE; PS00915; PI3 4 KINASE 1.
DR PROSITE; PS00916; PI3 4 KINASE 2.
KW TRANSFERASE; KINASE; CELL CYCLE.
FT DOMAIN 441 447 ARG/LYS-RICH (BASIC).
FT MOTAGEN 1972 1972 S->R,N: LOSS OF RAPAMYCIN SENSITIVITY.
FT CONFLICT 58 58 G->D (IN REF. 2).
FT CONFLICT 115 115 I->V (IN REF. 2).
FT CONFLICT 133 133 N->S (IN REF. 2).
FT CONFLICT 231 231 A->R (IN REF. 2).
FT CONFLICT 675 675 T->I (IN REF. 2).
FT CONFLICT 1292 1292 G->E (IN REF. 2).
FT CONFLICT 1436 1436 G->A (IN REF. 2).

FT CONFLICT 1469 1471 WGL -> GGS (IN REF. 2).
FT CONFLICT 1590 1590 V -> I (IN REF. 2).
FT CONFLICT 1632 1642 NDPSLPNTVKA -> TLIVYQIRSKP (IN REF. 2).
FT CONFLICT 1844 1844 L -> S (IN REF. 2).
FT CONFLICT 2202 2202 Q -> H (IN REF. 2).
SQ SEQUENCE 2470 AA; 281162 MW; 21744712 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXX

RVAVLWHELWYEGLEDASRQFFVEHNE
1960 X 1970 X 1980

7. US-08-121-713B-24 (1-8)

POLG EMCVB GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE
ID POLG EMCVB STANDARD; PRT; 2292 AA.
AC P17593;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE PROTEINS P2A TO
DE P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C (EC 3.4.22.28)
DE (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D (EC 2.7.7.48)).
OS ENCEPHALOMYOCARDITIS VIRUS (STRAIN EMC-B NONDIABETOGENIC).
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; PICORNAVIRIDAE; CARDIOVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RM 89243189
RA BAE Y.S., EUN H.M., YOON J.W.;
RL VIROLOGY 170:282-287(1989).
CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.
CC -1- PTM: SPECIFIC ENZYMOLOGIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
DR EMBL; M22457; EMCBCG.
DR PIR; B31473; GNVYEB.
DR HSP; P12296; IMEC.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; RNA-DIRECTED RNA POLYMERASE;
KW HYDROLASE; THIOLESTERASE; MYRISTYLATION.
FT PROPEP 1 67 LEADER PEPTIDE.
FT CHAIN 68 137 COAT PROTEIN VP4 (RHO).
FT CHAIN 138 393 COAT PROTEIN VP2 (BETA).
FT CHAIN 394 624 COAT PROTEIN VP3 (GAMMA).
FT CHAIN 625 901 COAT PROTEIN VP1 (ALPHA).
FT CHAIN 902 1038 CORE PROTEIN P2A (G).
FT CHAIN 1059 1194 CORE PROTEIN P2B (I).
FT CHAIN 1195 1519 CORE PROTEIN P2C (F).
FT CHAIN 1520 1607 CORE PROTEIN P3A.
FT CHAIN 1608 1627 GENOME-LINKED PROTEIN VPG (H).
FT CHAIN 1628 1832 PICORNAIN 3C (P22).
FT CHAIN 1833 2292 RNA-DIRECTED RNA POLYMERASE P3D (E).
FT CHAIN 68 68 MYRISTATE (BY SIMILARITY).
FT LIPID

FT ACT_SITE 1786 1786 PROTEASE (POTENTIAL).
FT ACT_SITE 1804 1804 PROTEASE (POTENTIAL).
SQ SEQUENCE 2292 AA; 255495 MW; 19494323 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXX

DAPLPCFQNNCLFLEKAGLQFRDNRKE
1470 1480

8. US-08-121-713B-24 (1-8)
POLG_EMCVD GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE

ID POLG_EMCVD STANDARD; PRT; 2292 AA.

AC P17594;
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE PROTEINS P2A TO
DE P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C (EC 3.4.22.28)
DE (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D (EC 2.7.7.48)).
OS ENCEPHALOMYOCARDITIS VIRUS (STRAIN EMC-D DIABETOGENIC).
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; PICORNAVIRIDAE; CARDIOVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RM 89243189
RA BAE Y.S., EUN H.M., YOON J.W.;
RA RIVIROLOGY 170:282-287(1989).
RL FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -1- PTM: SPECIFIC ENZYMTIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
DR EMBL; M22458; EMCDDG.
DR PIR; A31473; GNNYED.
DR HSP; P12296; IMEC.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; RNA-DIRECTED RNA POLYMERASE;
KW HYDROLASE; THIOLE PROTEASE; MYRISTYLATION.
FT PROPEP 1 67
FT CHAIN 68 137
FT CHAIN 138 393
FT CHAIN 394 624
FT CHAIN 625 901
FT CHAIN 902 1038
FT CHAIN 1059 1194
FT CHAIN 1195 1519
FT CHAIN 1520 1607
FT CHAIN 1608 1627
FT CHAIN 1628 1832
FT CHAIN 1833 2292
FT LIPID 68 68
FT ACT_SITE 1786 1786 RNA-DIRECTED RNA POLYMERASE P3D (E).
FT ACT_SITE 1804 1804 MYRISTATE (BY SIMILARITY).
FT ACT_SITE 1804 1804 PROTEASE (POTENTIAL).
SQ SEQUENCE 2292 AA; 255382 MW; 19455620 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXX

DAPLPCFQNNCLFLEKAGLQFRDNRKE
1470 1480

9. US-08-121-713B-24 (1-8)
POLG_EMCV GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE

ID POLG_EMCV STANDARD; PRT; 2290 AA.

AC P03304;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE PROTEINS P2A TO
DE P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C (EC 3.4.22.28)
DE (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D (EC 2.7.7.48)).
OS ENCEPHALOMYOCARDITIS VIRUS.
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; PICORNAVIRIDAE; CARDIOVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RM 84169586
RA PALMENBERG A.C., KIRBY E.M., JANDA M.R., DRAKE N.L., DUKE G.M.,
RA POTRAIZ K.F., COLLETT M.S.;
RL NUCLEIC ACIDS RES. 12:2969-2985(1984).
CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -1- PTM: SPECIFIC ENZYMTIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
DR EMBL; X00463; PIEMCY.
DR PIR; A03906; GNNVE.
DR HSP; P12296; IMEC.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; RNA-DIRECTED RNA POLYMERASE;
KW HYDROLASE; THIOLE PROTEASE; MYRISTYLATION.
FT PROPEP 1 67
FT CHAIN 68 136
FT CHAIN 137 391
FT CHAIN 392 622
FT CHAIN 623 910
FT CHAIN 911 1056
FT CHAIN 1057 1192
FT CHAIN 1193 1517
FT CHAIN 1518 1605
FT CHAIN 1606 1625
FT CHAIN 1626 1830
FT CHAIN 1831 2290
FT LIPID 68 68
FT ACT_SITE 1784 1784 MYRISTATE (BY SIMILARITY).
FT ACT_SITE 1802 1802 PROTEASE (POTENTIAL).
SQ SEQUENCE 2290 AA; 255756 MW; 19481048 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42

Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXX

EAPLPCFQNNCLFLEKAGLQFDRNRKE
1460 1470 1480

10. US-08-121-713B-24 (1-8)
POL1_GFLV RNA1 POLYPROTEIN (253 KD PROTEIN) (CONTAINS: 63 KD
POL1_GFLV STANDARD; PRT; 2284 AA.
ID AC P29149;
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DE 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE RNA1 POLYPROTEIN (253 KD PROTEIN) (CONTAINS: 63 KD PROTEASE COFACTOR,
72 KD MEMBRANE-BINDING PROTEIN, GENOME-LINKED PROTEIN (VEG), PROTEASE
(EC 3.4.22.-), RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48)).
DE OS GRAPEVINE FANLEAF VIRUS (GFLV).
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; NEPOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RM 92013951
RA RITZENTHALER C., VIRY M., PINCK M., MARGIS R., FUCHS M., PINCK L.;
RL J. GEN. VIROL. 72:2357-2365(1991).
CC -I- PFM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
DR EMBL; D00915; GFLRNA1.
DR PIR; J01373; GNVGV.
KW POLYPROTEIN; TRANSMEMBRANE; HYDROLASE; PROTEASE; TRANSFERASE;
RN RNA-DIRECTED RNA POLYMERASE; ATP-BINDING.
FT CHAIN 1 568 63 KD PROTEASE COFACTOR (POTENTIAL).
FT CHAIN 569 1217 72 KD MEMBRANE-BINDING PROTEIN
(POTENTIAL).
FT CHAIN 1218 1241 GENOME-LINKED PROTEIN (POTENTIAL).
FT CHAIN 1242 1460 CYSTEINE PROTEASE (POTENTIAL).
FT CHAIN 1461 2284 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT NP_BIND 781 788 ATP (POTENTIAL).
FT ACT_SITE 1420 1420 CYSTEINE PROTEASE (POTENTIAL).
SQ SEQUENCE 2284 AA; 252930 MW; 20307487 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXXX

ISPWVAVKCCVARLEDGIPQFHWSKYA
1360 1370 X 1380

11. US-08-121-713B-24 (1-8)
YCS2_YEAST HYPOTHETICAL 251.0 KD PROTEIN IN CRV1-RBK1 INTERGE

ID YCS2_YEAST STANDARD; PRT; 2167 AA.
AC P25356;
DT 01-MAY-1992 (REL. 22, CREATED)

DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 251.0 KD PROTEIN IN CRV1-RBK1 INTERGENIC REGION.
GN YCR32W OR YCR591 OR YCR601.
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RM 92133166
RA WICKSTEED B.L., ROBERTS A.B., SAGLIOCCO F.A., BROWN A.J.P.;
RL YEAST 7:761-772(1991).
RN [2]
RP SEQUENCE FROM N.A.
RM 92116648
RA RODRIGUEZ F., MARTEGANI E., MAURI I., ALBERGHINA L.;
RL YEAST 7:631-641(1991).
RN [3]
RP SEQUENCE OF 313-2167 FROM N.A.
RM 91335897
RA JIA Y., SLONIMSKI P.P., HERBERT C.J.;
RL YEAST 7:413-424(1991).
CC -I- THIS PROTEIN IS ENCODED BY A NON-ESSENTIAL GENE.
DR EMBL; X59075; SCYCR59.
DR EMBL; X59720; SCCHRIII.
DR PIR; S15052; S15052.
DR PIR; S19444; S19444.
DR LISTA; SC01404; YCR32W.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 2167 AA; 250972 MW; 19698864 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X
XXXXXXX

TSNRMTIEKHIVLEIKGQFCISNDNY
260 X 270

12. US-08-121-713B-24 (1-8)
MYSN_HUMAN MYOSIN HEAVY CHAIN, NONMUSCLE TYPE A (CELLULAR MYO

ID MYSN_HUMAN STANDARD; PRT; 1961 AA.
AC P35579;

DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE MYOSIN HEAVY CHAIN, NONMUSCLE TYPE A (CELLULAR MYOSIN HEAVY CHAIN,
DE TYPE A) (NMHC-A).
GN MYH9.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE OF 1-1337 FROM N.A.
RM 92003925

RA TOOTHAKER L.E., GONZALEZ D.A., TUNG N., LEMONS R.S., LE BEAU M.M.,
RA ARNAOUT M.A., CLAYTON L.K., TENEN D.G.;

RL BLOOD 78:1826-1833(1991).
RN [2]
RP SEQUENCE OF 1-715 FROM N.A.
RM 91316803
RA STOMOS M., WANG M., MCBRIDE O.W., KAWAMOTO S., YAMAKAWA K.,
RA GOLA D., ADELSTEIN R.S., WEIR L.;
RL CIRC. RES. 69:530-539(1991).
[3]
RN SEQUENCE OF 715-1961 FROM N.A.
RM 90138958
RA SAEZ C.G., MYERS J.C., SHOWS T.B., LEINWAND L.A.;
RL PROC. NATL. ACAD. SCI. U.S.A. 87:1164-1168(1990).
CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC CAPPING.
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (IMM)
CC AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATTER BE SPLIT FURTHER INTO
CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY. STRONGEST
CC TO OTHER NONMUSCLE MYOSINS.
DR EMBL; M81105; HSMYH9.
DR EMBL; M89180; HSMYOHCA.
DR EMBL; M31013; HSMYOHCA.
DR HSSP; P24733; 1SCM.
DR MIM; 160775; 11TH EDITION.
KW MYOSIN; COILED COIL; ACTIN-BINDING; ALKYLATION; ATP-BINDING;
KW HEPTAD REPEAT PATTERN; MULTIGENE FAMILY.
FT DOMAIN 1 835 GLOBULAR HEAD (S1).
FT DOMAIN 836 1926 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT NP BIND 174 181 ATP.
FT DOMAIN 654 676 ACTIN-BINDING.
FT MOD RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 53 55 EAL -> RGH (IN REF. 2).
FT CONFLICT 660 660 T -> S (IN REF. 2).
FT CONFLICT 869 869 T -> M (IN REF. 3).
FT CONFLICT 931 931 C -> Y (IN REF. 3).
FT CONFLICT 1240 1241 KG -> GR (IN REF. 3).
SQ SEQUENCE 1961 AA; 226600 MW; 15612344 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X X
XXXXXXX
ARALEAEKAELERLNKKQFTMEDL
1490 X 1500

13. US-08-121-713B-24 (1-8)
MYSN_CHICK MYOSIN HEAVY CHAIN, NONMUSCLE (NMHMC).

ID MYSN_CHICK STANDARD; PRT; 1959 AA.
AC P14105;
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE MYOSIN HEAVY CHAIN, NONMUSCLE (NMHMC).
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINAL EPITHELIUM;
RM 90046668
RA SHORST R.V., CONTI M.A., KAWAMOTO S., PRESTON Y.A., BRILL D.A.,
RA ADELSTEIN R.S.;
RL PROC. NATL. ACAD. SCI. U.S.A. 86:7726-7730(1989).
CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC CAPPING.
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (IMM)
CC AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATTER BE SPLIT FURTHER INTO
CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY. STRONGEST
CC TO OTHER NONMUSCLE MYOSINS.
DR EMBL; M26510; GGMVHN.
DR PIR; A33977; A33977.
DR HSSP; P24733; 1SCM.
KW MYOSIN; COILED COIL; ACTIN-BINDING; ALKYLATION; ATP-BINDING;
KW HEPTAD REPEAT PATTERN; MULTIGENE FAMILY.
FT DOMAIN 1 835 GLOBULAR HEAD (S1).
FT DOMAIN 836 1926 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT NP BIND 174 181 ATP.
FT DOMAIN 654 676 ACTIN-BINDING.
FT MOD RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1959 AA; 226502 MW; 15394260 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0

X X X
XXXXXXX
ARALEAEKAELERLNKKQFTMEDL
1490 X 1500

14. US-08-121-713B-24 (1-8)
TRHY_HUMAN TRICHOHYALIN.

ID TRHY_HUMAN STANDARD; PRT; 1898 AA.
AC Q07283;
DT 01-OCT-1994 (REL. 30, CREATED)

DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE TRICHOVALIN.
GN THH OR TRHY OR THL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
[1]
RN SEQUENCE FROM N.A.
RM 93280194
RA LEE S.-C., KIM I.-G., MAREKOV L.N., O'KEEFE E.J., PARRY D.A.D.,
RA STEINERT P.M.;
RL J. BIOL. CHEM. 268:12164-12176(1993).
[2]
RN SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RM 93315897
RA O'KEEFE E.J., HAMILTON E.H., LEE S.-C., STEINERT P.M.;
RL J. INVEST. DERMATOL. 101:658-715(1993).
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD, PROTEIN, TOGETHER
WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
DIFFERENTIATION.
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
THE EPIDERMIS.
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
DIFFERENT SPECIES.
CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS, AND
MORE SPECIFICALLY TO S-100/CABP LIKE PROTEINS IN THE N-TERMINUS.
DR EMBL; L09190; HSTRHYAL.
DR PIR; A45973; A45973.
DR HSP; P02633; IBOC.
DR MIM; 190370; 11TH EDITION.
KW REPEAT; CALCIUM-BINDING.
FT DOMAIN 1 91 S-100 LIKE.
FT CA BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).
FT CA BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
FT DOMAIN 314 390 6 X 13 AA TANDEM REPEATS OF
R-R-E-Q-E-E-F-R-R-E-Q-Q-L.
FT REPEAT 314 326 1-1 (APPROXIMATE).
FT REPEAT 327 339 1-2 (APPROXIMATE).
FT REPEAT 340 351 1-3 (APPROXIMATE).
FT REPEAT 352 364 1-4.

FT REPEAT 365 377 1-5.
FT REPEAT 378 390 1-6.
FT DOMAIN 391 444 9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
FT REPEAT 391 396 2-1.
FT REPEAT 397 402 2-2.
FT REPEAT 403 408 2-3.
FT REPEAT 409 414 2-4.
FT REPEAT 415 420 2-5.
FT REPEAT 421 426 2-6.
FT REPEAT 427 432 2-7.
FT REPEAT 433 438 2-8.
FT REPEAT 439 444 2-9.
FT DOMAIN 444 702 9 X 28 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 923 1162 8 X 30 AA TANDEM REPEATS.
FT REPEAT 923 952 4-1.
FT REPEAT 953 982 4-2.
FT REPEAT 983 1012 4-3.
FT REPEAT 1013 1042 4-4.
FT REPEAT 1043 1072 4-5.
FT REPEAT 1073 1102 4-6.
FT REPEAT 1103 1132 4-7.
FT REPEAT 1133 1162 4-8.
FT DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT CONFLICT 1752 1752 F -> L (IN REF. 2).
FT CONFLICT 1794 1801 OERDQYR -> RSETGSTG (IN REF. 2).
FT CONFLICT 1857 1857 Q -> K (IN REF. 2).
FT CONFLICT 1880 1880 V -> G (IN REF. 2).
SQ SEQUENCE 1898 AA; 247219 MW; 12293610 CN;
Initial Score = 8 Optimized Score = 8 Significance = 7.42
Residue Identity = 0% Matches = 0 Mismatches = 8
Gaps = 0 Conservative Substitutions = 0
X X
XXXXXXX
EQSQEGKGHGRLPEGTHQFASVFRS
1860 X 1870 X 1880
15. US-08-121-713B-24 (1-8)
MSAS_PENPA 6-METHYLSALICYLIC ACID SYNTHASE (EC 2.3.1.1-) (MSAS)
ID MSAS_PENPA STANDARD; PRT; 1774 AA.
AC P22367.
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE 6-METHYLSALICYLIC ACID SYNTHASE (EC 2.3.1.1-) (MSAS).
OS PENICILLIUM PATULUM (PENICILLIUM GRISEOFULVUM).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; PLECTOMYCETES; EUROTIALES.
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN=DSM 62862;
RM 91006137
RA BECK J., RIPKA S., SIEGNER A., SCHILTZ E., SCHWEIZER E.;
RL EUR. J. BIOCHEM. 192:487-498(1990).
CC -1- FUNCTION: THIS MULTIFUNCTIONAL ENZYME IS A POLYKETIDE SYNTHASE.
IT CATALYSES A TOTAL OF 11 STEPS BY SEVEN DIFFERENT COMPONENT
CC ENZYMES, IN THE BIOSYNTHESIS OF THE ANTIBIOTIC PATULIN.

CC -!- CATALYTIC ACTIVITY: ACETYL-COA + 3 MALONYL-COA + NADPH =
 CC 6-METHYLSALICYLIC ACID + NADP(+) + 3 COA + 3 CO(2) + H(2)O.
 CC -!- PATHWAY: BIOSYNTHESIS OF PATULIN.
 CC -!- SUBUNIT: HOMOMULTIMER.
 CC -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
 CC -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
 CC FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
 CC THIOLASES.
 CC EMBL; X55776; PPMAS.
 DR PIR; S13178.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE.
 DR PROSITE; PS00606; B KETOACYL SYNTHASE.
 KW MULTIFUNCTIONAL ENZYME; OXIDOREDUCTASE; ANTIBIOTIC BIOSYNTHESIS;
 KW TRANSFERASE; NADP; PHOSPHOPANTETHEINE.
 FT DOMAIN 186 238 ACYLTRANSFERASE (AT).
 FT DOMAIN 642 676 ACETYL/MALONYL TRANSFERASES.
 FT DOMAIN 1403 1450 2-OXOACYL REDUCTASES.
 FT DOMAIN 1697 1758 2-OXOACYL SYNTHASE.
 FT DOMAIN 1697 1758 ACYL CARRIER.
 FT NP BIND 1419 1424 NADP (POTENTIAL).
 FT ACT_SITE 204 204 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT ACT_SITE 653 653 MALONYLTRANSFERASE (BY SIMILARITY).
 FT BINDING 1732 1732 PHOSPHOPANTETHEINE (BY SIMILARITY).
 SQ SEQUENCE 1774 AA; 190732 MW; 15502921 CN;

Initial Score = 8 Optimized Score = 8 Significance = 7.42
 Residue Identity = 0% Matches = 0 Mismatches = 8
 Gaps = 0 Conservative Substitutions = 0

X X
 XXXXXXXX

LKNITSRGYFLDRLEDFDCQFFGISPK
 90 100 X 110

IntelliGenetics

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq30asq.res made by on Fri 19 May 95 8:42:39-PDT.

Query sequence being compared:US-08-121-713B-30 (1-7)
 Number of sequences searched: 53402
 Number of scores above cutoff: 3806

Results of the initial comparison of US-08-121-713B-30 (1-7) with:
Data bank : A-Geneseq 18, all entries

100000-
-
-
N
U50000*
-
M
B
E
R
C

SENDER

500-

— 100 —

*

05-1

10-

30

STDEV	SCORE
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	1
9	2
10	1

PARAMETERS

	2	20	6
Similarity matrix	1	1	1
Mismatch penalty	1.00	1.00	1.00
Gap penalty	0.05	0.05	0.05
Gap size penalty	0	0	0
Cutoff score	0	0	0
Randomization group	0	0	0
Unitary	1	1	1
K-tuple	2	20	6
Joining penalty	1	1	1
Window size	1	1	1

Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	0	1	1.23

Times:	CPU	Total Elapsed
	00:00:31.12	00:00:32.00

Number of residues: 6354270
Number of sequences searched: 53402
Number of scores above cutoff: 3806

Cut-off raised to 1.
Cut-off raised to 3.
Cut-off raised to 4.

The scores below are sorted by initial score. Significance is calculated based on initial score.

42 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
1. R13887	Inositol-3-phosphate binding	2749	7	7	5.71 0
2. R41043	CD4-EBA175 fusion protein.	1786	7	7	5.71 0
3. R36780	KRE5.	1365	7	7	5.71 0
4. P94265	Sequence of APH36.1 clone.	783	7	7	5.71 0
5. R33424	G6PD.	751	7	7	5.71 0
6. R41334	91 kD ISGF-3alpha.	739	7	7	5.71 0
7. R41335	84 kD ISGF-3alpha.	701	7	7	5.71 0
8. P70458	Sequence of gpb encoded by se	484	7	7	5.71 0
9. R25930	Human liver Type I Iodothyron	249	7	7	5.71 0
10. R44510	Type I Iodothyronine 5' deiod	249	7	7	5.71 0

11. R33383	Cytokine mg-CSF.	208	7	7	5.71	0
12. P90051	Human granulocyte colony stim	207	7	7	5.71	0
13. P70161	Human G-CSF encoded by pBRG4	207	7	7	5.71	0
14. P70163	Human G-CSF deduced from geno	207	7	7	5.71	0
15. P71383	Sequence of human granulocyte	207	7	7	5.71	0
16. P61340	Plasmid pBRG4 granulocyte CSF	207	7	7	5.71	0
17. P71384	Sequence of human granulocyte	204	7	7	5.71	0
18. P70731	Sequence encoded by human gra	204	7	7	5.71	0
19. P61341	Plasmid pBRV2 insert encoding	204	7	7	5.71	0
20. R33382	Cytokine hG-CSF.	204	7	7	5.71	0
21. P95033	Granulocyte colony stimulat	204	7	7	5.71	0
22. P70162	Human G-CSF encoded by pBRV2	204	7	7	5.71	0
23. P83163	Sequence of human granulocyte	204	7	7	5.71	0
24. P70730	Sequence encoded by human gra	186	7	7	5.71	0
25. R59511	Sequence of the light chain v	133	7	7	5.71	0
26. R10920	kappa light chain variable re	132	7	7	5.71	0
27. R37716	Mouse 4C10 anti-idiotypic Ab l	132	7	7	5.71	0
28. P90543	Amino acids sequence of a v c	131	7	7	5.71	0
29. R04132	Anti-leu 3a light chain varia	131	7	7	5.71	0
30. R05089	Light chain variable domain o	131	7	7	5.71	0
31. R28668	p12-k2.	131	7	7	5.71	0
32. R29008	p64-k4 protein product.	131	7	7	5.71	0
33. R30881	Antibody 4A2 light chain cons	131	7	7	5.71	0
34. R32123	Anti-CD4 antibody MT 3.10 lig	131	7	7	5.71	0
35. R06375	Anti-Tac light chain variable	126	7	7	5.71	0
36. R24108	Humanised anti-Tac antibody l	126	7	7	5.71	0
37. R04938	Chelate-specific light chain	110	7	7	5.71	0
38. R05038	Metal chelate specific light	110	7	7	5.71	0
39. R94136	CH255 light chain variable r	110	7	7	5.71	0
40. R10540	Eukaryotic signal peptide enc	20	7	7	5.71	0
41. R54141	CH255 light chain variable r	9	7	7	5.71	0
42. R65508	Endothelin antagonist cyclic	5	7	7	5.71	0

The list of other best scores is:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
43. P50687	Hepatitis B virus antigenic p	16	4	4	3.26
44. R37946	HCV NS-4 type 1 region 2 (171	17	4	4	3.26
45. R26013	Influenza fusion peptide #10.	22	4	4	3.26
1. US-08-121-713B-30 (1-7)	Inositol-3-phosphate binding peptide.				
RI3887					
ID	RI3887 standard; Protein; 2749 AA.				
AC	RI3887;				
DT	27-NOV-1991 (first entry)				
DE	Inositol-3-phosphate binding peptide.				
KW	I-3-P; transformation; inositol.				
PN	J03183482-A.				
PD	09-AUG-1991.				
PF	14-DEC-1989; JP-324256.				
PR	14-DEC-1989; JP-324256.				
PA	(KYOW) KYOWA HAKKO KOGYO KK.				
DR	WPI; 91-277584/38.				
DR	N-PSDB; Q13593.				
PT	New polypeptide having binding affinity to inositol-3-phosphate -				

PT	prepd. by culturing cell contg. recombinant plasmid comprising
PT	DNA and vector DNA
PS	Disclosure; Fig 2(1-3); 11pp; Japanese.
CC	The sequence encoding this peptide may be included in a
CC	plasmid/vector for transformation of a host cell and mass-prodn.
CC	of the peptide.
SQ	Sequence 2749 AA;
SQ	146A; 155R; 144N; 154D; 0 B; 60 C; 128Q; 229E; 0 Z; 142G; 71 H;
SQ	155T; 314I; 167K; 69 M; 123F; 99 P; 180S; 136T; 21 W; 71 Y; 185V;
Initial Score	= 7 Optimized Score = 7 Significance = 5.71
Residue Identity	= 28% Matches = 2 Mismatches = 5
Gaps	= 0 Conservative Substitutions = 0
	X X X
	XWXXXIX
	GTLEPHWSGLLWTAMLSIAIVIALPK
	2310 X 2320
2. US-08-121-713B-30 (1-7)	CD4-EBA175 fusion protein.
R41043	
ID	R41043 standard; protein; 1786 AA.
AC	R41043;
DT	22-MAR-1994 (first entry)
DE	CD4-EBA175 fusion protein.
KW	Merozoite; Erythrocyte Binding Antigen 175; malaria; HIV; env;
KW	human immunodeficiency virus; envelope glycoprotein; hybrid protein;
KW	red blood cell; erythrocyte; AIDS; molecular machine.
OS	Chimeric Homo sapiens.
OS	Chimeric Plasmodium falciparum.
FH	Key Location/Qualifiers
FT	Region 1..371
FT	/note= "residues 1-371 of CD4"
FT	Region 372..1786
FT	/note= "residues 20-1435 of EBA-175"
PN	W09318160-A.
PD	16-SEP-1993.
PF	10-MAR-1993; G00505.
PR	11-MAR-1992; GB-005276.
PR	08-JUL-1992; GB-014481.
PR	24-JUL-1992; GB-015829.
PR	16-SEP-1992; GB-019562.
PR	03-MAR-1993; GB-004511.
PA	(PREN/) PRENDERGAST K F.
PI	Prendergast KF;
DR	WPI; 93-303474/38.
PT	Anti-viral fusion peptide(s) - comprise viral-binding component
PT	and malaria merozoite red cell binding component, for treating
PT	e.g. HIV, and hepatitis
PS	Claim 9; Page 44-47; 69pp; English.
CC	The hybrid protein NH2-CD4(1-371)-EBA175(20-1435)-COOH is a
CC	specifically claimed example of a fusion protein of the invention;
CC	it comprises at least part of the CD4 molecule fused to a peptide
CC	from a malarial parasite merozoite protein with affinity for red
CC	blood cells. The fusion protein can bind free HIV in the blood to
CC	red blood cells and consequently reduce viral titre, prevent
CC	transmission of the virus and improve safety of blood transfusions.

SQ Sequence 1786 AA;
 SQ 47 A; 65 R; 163N; 126D; 0 B; 41 C; 63 Q; 169E; 0 Z; 73 G; 40 H;
 SQ 103I; 134L; 202K; 29 M; 55 F; 51 P; 160S; 98 T; 28 W; 54 Y; 85 V;
 Initial Score = 7 Optimized Score = 7 Significance = 5.71
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X X
 X XXXXXLX

KEWNEFREKLWEAMLSHEKNNINCK
 650 X 660

3. US-08-121-713B-30 (1-7)
 R36780 KRE5.

ID R36780 standard; Protein; 1365 AA.
 AC R36780;
 DT 16-JUL-1993 (first entry)
 DE KRE5.

KW Yeast; cell wall; beta-glucan; assembly; pathway; KRE1; KRE5; growth;
 KW secretory; O-linked mannose; (>6)-beta-glucan; epistasis; morphology;
 KW hydrophilic; glycoprotein; COOH-terminal; endoplasmic reticulum; ER;
 KW retention signal; antifungal agent.
 OS Saccharomyces cerevisiae.
 PN US5194600-A.
 PD 16-MAR-1993.

PF 05-MAR-1990; 488316.
 PR 05-MAR-1990; US-488316.
 PA (ROYA-) ROYAL INST ADVANCEMENT LEARNING.
 PI Boone C, Bussey H, Hill K, Meaden P, Sommer SS;
 DR WPI; 93-109384/13.
 DR N-PSDB; Q38899.

PT New DNA encoding genes which participate in beta-glucan assembly
 PT - useful for producing mutants for in-vivo screening of
 PT antifungal agents and providing tools for in-vitro screening
 PS Claim 1; Columns 38-44; 24pp; English.
 CC The sequences given in R34785 and R36780 represent proteins which
 CC participate in a yeast cell wall beta-glucan assembly pathway.
 CC These proteins represent KRE1 and KRE5 respectively, and are
 CC essential for normal cell growth. KRE1 is a Ser/Thr rich protein
 CC that is directed into the yeast secretory pathway, where it is
 CC highly modified, probably through addition of O-linked mannose
 CC residues. Gene disruption of the KRE1 locus leads to a 40% reduced
 CC level of cell wall (>6)-beta-glucan. Mutations at KRE5 also caused
 CC defects in cell wall (>6)-beta-glucan production and appears to be
 CC epistatic to KRE1. KRE5 is a large hydrophilic secretory glyco-
 CC protein which contains the COOH-terminal endoplasmic reticulum (ER)
 CC retention signal (His-Ap-Glu-Leu). Deletion of the KRE5 gene results
 CC in cells with aberrant morphology and extremely compromised growth.
 CC KRE1 and KRE5 are useful as tools for the in vitro screening of anti-
 CC fungal agents which inhibit fungi pathogenic to plants and animals.
 CC The genes can be used to produce mutants for in vivo screening of
 CC antifungal agents.

SQ Sequence 1365 AA;
 SQ 43 A; 56 R; 81 N; 98 D; 0 B; 12 C; 51 Q; 92 E; 0 Z; 60 G; 23 H;
 SQ 114I; 166L; 92 K; 25 M; 66 F; 60 P; 121S; 61 T; 13 W; 55 Y; 76 V;

Initial Score = 7 Optimized Score = 7 Significance = 5.71
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X X
 X XXXXXLX

MSGDGYWKEGYWKMRLRENNLEFYSTE
 1230 1240 1250

4. US-08-121-713B-30 (1-7)
 P94265 Sequence of APH36.1 clone.

ID P94265 standard; protein; 783 AA.
 AC P94265;
 DT 24-JUN-1990 (first entry)
 DE Sequence of APH36.1 clone.
 KW Clone APH36.1; acyl-peptide hydrolase.

OS Rattus rattus.
 FH Key Location/Qualifiers
 FT Protein 1..721
 FT /label=claimed protein
 FT Active-site 580..587
 FT Region 130.
 FT /note="potential site of glycosylation"
 FT Region 131.
 FT /note="potential site of glycosylation"
 FT Region 132.
 FT /note="potential site of glycosylation"
 FT Region 229.
 FT /note="potential site of glycosylation"
 FT Region 230.
 FT /note="potential site of glycosylation"
 FT Region 231.
 FT /note="potential site of glycosylation"
 FT Region 239.
 FT /note="potential site of glycosylation"
 FT Region 240.
 FT /note="potential site of glycosylation"
 FT Region 241.
 FT /note="potential site of glycosylation"
 PN EP-303997-A.
 PD 22-FEB-1989.
 PF 13-AUG-1988; 113186.
 PR 21-AUG-1987; US-087936.
 PA (GEO-) The General Hospital Corp.
 PI Smith JA;
 DR WPI; 89-055444/08.
 DR N-PSDB; N91052.

PT Recombinant acyl-peptide hydrolase -used to catalyse hydrolysis of N-acyl
 PT peptide(s) or reaction of N-acetyl amino acid donor and acceptor protein.
 PS Disclosure; pp; English.
 CC Sequence contains the claimed sequence of acyl-peptide hydrolase (APH)
 CC (claim 1, page 11). APH can be used to catalyse the hydrolysis of the
 CC N-terminal acyl amino acid of an acylated polypeptide, or the reaction
 CC between a derivatised N alpha-acetyl amino acid donor and acceptor with
 CC a free alpha-NH2. It can also be used to make refractory proteins
 CC susceptible to Edman sequencing or th reduce degradation of proteins to
 CC be used therapeutically.

SQ Sequence 783 AA; 40 D; 0 B; 20 C; 37 Q; 44 E; 0 Z; 57 G; 19 H; 50 A; 39 R; 18 N; 21 M; 31 F; 53 P; 73 S; 33 T; 19 W; 24 Y; 61 V; 24 I; 81 L; 39 K; 21 M; 31 F; 53 P; 73 S; 33 T; 19 W; 24 Y; 61 V;
Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX

YNSCLPDLNVWEMLDKSPIKIPQV
640 X 650

5. US-08-121-713B-30 (1-7)
R33424 G6PD.

ID R33424 standard; Protein; 751 AA.
AC R33424;
DT 27-JUL-1993 (first entry)
DE G6PD.
KW Polymerase chain reaction; glucose-6-phosphate dehydrogenase; PCR;
E. coli; G6PD; drug; transformation; primer; amplify.
ON Plasmodium falciparum.
PS WO9306125-A.
PD 01-APR-1993.
PF 21-SEP-1992; U07807.
PR 20-SEP-1991; US-782137.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Kaslow DC, Shahabuddin M;
DR WPI; 93-117467/14.
DR N-PSDB; Q33424.
PT DNA segments encoding Plasmodium falciparum G6PD obtd. by PCR -
used to express proteins and raise antibodies for diagnosis and
treatment of malaria
PT Claim 2; Fig 2; 36pp; English.

CC This sequence represents Plasmodium falciparum glucose-6-phosphate
dehydrogenase (G6PD). The DNA encoding this sequence was isolated
using the primers given in Q38806-07. The amplified sequence was
used in the production of transformed E. coli which produce a
recombinant P. falciparum G6PD. These transformed cells can be used
in a method of screening drugs for activity against P. falciparum
G6PD.
SQ Sequence 751 AA; 44 D; 0 B; 18 C; 17 Q; 52 E; 0 Z; 26 G; 11 H; 15 A; 14 R; 85 N; 44 D; 0 B; 18 C; 17 Q; 52 E; 0 Z; 26 G; 11 H; 69 I; 73 L; 80 K; 15 M; 41 F; 22 P; 50 S; 36 T; 4 W; 53 Y; 26 V;
Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX

FLNSKDLKLDLWKNLIKSYIEVNYNL
80 90 X 100

6. US-08-121-713B-30 (1-7)

R41334 91 kD ISGF-3alpha.

ID R41334 standard; Protein; 739 AA.
AC R41334;
DT 22-APR-1994 (first entry)
DE 91 kD ISGF-3alpha.
KW 113 kD; 91 kD; 84 kD; ISGF-3alpha; interferon-related; receptor;
DNA binding factor; gene family; translation protein; tyrosine;
DNA binding protein; interferon-gamma; hairy cell leukaemia;
interferon therapy; chronic viral hepatitis; phosphorylation;
adjuvant therapy; tyrosine kinase.
OS Homo sapiens.
PN WO9319179-A.
PD 30-SEP-1993.
PF 19-MAR-1993; U02569.
PR 19-MAR-1992; US-854296.
PR 23-NOV-1992; US-980498.
PA (UYRQ) UNIV ROCKEFELLER.
PI Darnell JE, Fu X, Schindler CW, Shuai K;
DR N-PSDB; Q49165.
PT Interferon receptor recognition factors - useful e.g. to treat
viral hepatitis, hairy cell leukaemia and to potentiate interferon
effects
PS Claim 17; Fig 2; 131pp; English.

CC The sequences given in R41333-35 represent the 113 kD, 91 kD and 84 kD
ISGF-3alpha proteins respectively. ISGF-3alpha is an interferon-
related receptor recognition factor which comprises several
substituents. The 113 kD, and the 91 and 84 kD proteins are derived
from two different but related genes. It is clear that a gene
family exists and further members are likely to be found. The 91 kD
protein has the capability of acting as a translation protein and as
a DNA binding protein in response to interferon-gamma stimulation.
CC These proteins participate in rapid phosphorylation and dephosphory-
lation during the course of, and as part of their activity. This
phosphorylation takes place in an interferon-dependant manner on
specified tyrosine residues. These proteins may be used in
conjunction with interferon therapy eg. to treat chronic viral
hepatitis, hairy cell leukaemia and for use with interferon in
adjuvant therapy.
SQ Sequence 739 AA; 26 A; 34 R; 44 N; 38 D; 0 B; 10 C; 53 Q; 64 E; 0 Z; 27 G; 16 H; 36 I; 89 L; 55 K; 19 M; 33 F; 35 P; 49 S; 35 T; 14 W; 18 Y; 44 V;
Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX

SQLPGWASILWYNMLVAEPNLSFFL
460 470

7. US-08-121-713B-30 (1-7)
R41335 84 kD ISGF-3alpha.

ID R41335 standard; Protein; 701 AA.
AC R41335;

22-APR-1994 (first entry)
DE 84 kD ISGF-3alpha.
KW 113 kD; 91 kD; 84 kD; ISGF-3alpha; interferon-related; receptor;
KW recognition factor; gene family; translation protein; tyrosine;
KW DNA binding protein; interferon-gamma; hairy cell leukaemia;
KW interferon therapy; chronic viral hepatitis; phosphorylation;
KW adjuvant therapy; tyrosine kinase.
OS Homo sapiens.
PN WO9319179-A.
PD 30-SEP-1993.
PF 19-MAR-1993; U02569.
PR 19-MAR-1992; US-834296.
PR 23-NOV-1992; US-980498.
PA (GYRQ) UNIV ROCKEFELLER.
PI Darnell JE, Fu X, Schindler CW, Shuai K;
DR WPI; 93-320745/40.
DR N-PSDB; Q49166.
PT Interferon receptor recognition factors - useful e.g. to treat
PT viral hepatitis, hairy cell leukaemia and to potentiate interferon
PT effects
PS Claim 17; Fig 3; 131pp; English.
CC The sequences given in R41333-35 represent the 113 kD, 91 kD and 84 kD
CC ISGF-3alpha proteins respectively. ISGF-3alpha is an interferon-
CC related receptor recognition factor which comprises several
CC substituents. The 113 kD, and the 91 and 84 kD proteins are derived
CC from two different but related genes. It is clear that a gene
CC family exists and further members are likely to be found. The 91 kD
CC protein has the capability of acting as a translation protein and as
CC a DNA binding protein in response to interferon-gamma stimulation.
CC These proteins participate in rapid phosphorylation and dephosphory-
CC lation during the course of, and as part of their activity. This
CC phosphorylation takes place in an interferon-dependant manner on
CC specified tyrosine residues. These proteins may be used in
CC conjunction with interferon therapy eg. to treat chronic viral
CC hepatitis, hairy cell leukaemia and for use with interferon in
CC adjuvant therapy.
SQ Sequence 701 AA;
SQ 26 A; 33 R; 42 N; 35 D; 0 B; 10 C; 52 Q; 60 E; 0 Z; 26 G; 15 H;
SQ 35 I; 85 L; 55 K; 16 M; 31 F; 32 P; 44 S; 32 T; 14 W; 18 Y; 40 V;
Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX
|
SOLPSGWASIIWYNMLVAEPNLSFFL
460 470

8. US-08-121-713B-30 (1-7)
P70458 Sequence of gpD encoded by segment of Xanthomonas

ID P70458 standard; protein; 484 AA.
AC P70458;
DT 13-FEB-1991 (first entry)
DE Sequence of gpD encoded by segment of Xanthomonas campestris DNA
DE that contains a gene cluster that directs Xanthan biosynthesis.
KW Thickening agent; oil recovery; drilling fluid.

OS Xanthomonas campestris.
PN WO8705938-A.
PD 08-OCT-1987.
PF 24-MAR-1987; U00604.
PR 23-MAR-1987; US-029530.
PA (GETT-) Getty Sci Dev Co.
PI Capage MA, Doherty DH, Betlach MR, Vanderslice RW.
DR WPI; 87-291651/41.
PT Recombinant DNA prodn. of xanthan gum or its variants - by
PT transforming host cells with vector contg. DNA coding for enzymes
PT involved in polysaccharide synthesis
PS Example; Fig 12; 149pp; English.
CC Virtually all of the segment of Xanthomonas campestris DNA that
CC contains a gene cluster that directs Xanthan biosynthesis (N70753),
CC codes for protein products. Each gene is designated by a letter (see
CC Fig 11) and its protein product is designated by that letter
CC preceded by 'gp' (P70455-67).
SQ Sequence 484 AA;
SQ 33 A; 40 R; 15 N; 21 D; 0 B; 2 C; 23 Q; 13 E; 0 Z; 41 G; 12 H;
SQ 26 I; 61 L; 15 K; 13 M; 22 F; 21 P; 26 S; 21 T; 10 W; 19 Y; 50 V;
Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX
|
KILAVIALMGLMPLMIAIATVGVKMSSP
300 310 X 320

9. US-08-121-713B-30 (1-7)
R25930 Human liver Type I iodothyronine 5' deiodinase.

ID R25930 standard; Protein; 249 AA.
AC R25930;
DT 27-JAN-1993 (first entry)
DE Human liver Type I iodothyronine 5' deiodinase.
KW thyroid hormone; Secys; selenocysteine; thyroxine;
KW 3,3',5-triiodothyronine; ITDL.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Modified site 126
FT /label= OTHER
FT /note= "Selenocysteine"
PN WO9213077-A.
PD 06-AUG-1992.
PF 29-JAN-1992; U00740.
PR 29-JAN-1991; US-647657.
PR 03-SEP-1991; US-757024.
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
PI Berry MA, Larsen PR;
DR WPI; 92-284662/34.
DR N-PSDB; 026751.
PT Type I iodothyronine 5' deiodinase and DNA encoding it - for
PT treatment and diagnosis of thyroid function related diseases;
PT also for diagnosis of thyroid cancer
PS Claim 2; Fig 4B; 104pp; English.
CC A 2.1kb cDNA for 5'-deiodinase was isolated from a rat liver

CC library by expression in Xenopus oocytes (see Q26750). The clone
CC was used to probe a human cDNA library and a gene encoding the
CC complete human ITPI enzyme was isolated. The UGA codon at
CC nucleotide 382 encodes the rare amino acid selenocysteine. The
CC selenocysteine insertion sequence (SECIS) in the 3'-UTR has 66%
CC homology to the SECIS downstream of the rat ITPI gene. The
CC presence of SeCys is essential for full activity.
SQ Sequence 249 AA;
SQ 13 A; 16 R; 14 N; 10 D; 0 B; 4 C; 14 Q; 13 E; 0 Z; 14 G; 7 H;
SQ 12 I; 30 L; 11 K; 7 M; 15 F; 14 P; 13 S; 8 T; 8 W; 6 Y; 19 V;
SQ 1 Others;

Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXLX
|
MGLPQGLWLKRLWLVLEVAHVHV
10 X 20

10. US-08-121-713B-30 (1-7)
R44510 Type I iodothyronine 5' deiodinase.

ID R44510 standard; Protein; 249 AA.
AC R44510;
DT 16-JUN-1994 (first entry)
DE Type I iodothyronine 5' deiodinase.
KW Type I; iodothyronine 5'; deiodinase; 3' untranslated region;
KW selenocysteine; transient expression assay; antibody; diagnosis;
KW thyroid; function; selenocysteine insertion sequence; reporter gene;
KW transfection; efficiency; promoter.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc difference 126
FT /label= OTHER
FT /note= "Selenocysteine"
PN US5272078-A.
PD 21-DEC-1993.
PF 29-JAN-1991; 647657.
PR 29-JAN-1991; US-647657.
PR 03-SEP-1991; US-757024.
PR 29-JAN-1992; US-828790.
PA (SGHM) BRIGHAM & WOMENS HOSPITAL.
PI Berry MJ, Larsen PR;
DR N-PSDB; Q53466.
PT DNA encoding the type I iodo-thyronine 5'-deiodinase and mutants
PT of this - has a seleno-cysteine site which may be used to study
PT thyroid hormone and in diagnosis of thyroid cancer
PT Claim 2; Fig 4B; 49pp; English.
PS This sequence represents a type I iodothyronine 5' deiodinase. The
CC gene encoding this protein contains in the 3' untranslated region a
CC sequence which causes inclusion of a selenocysteine residue at a TGA
CC codon within the deiodinase gene. This protein and antibodies
CC reacting with it, are useful in the diagnosis and treatment of
CC disease states related to thyroid function. Mutant reporter genes
CC based on the deiodinase gene sequence containing Cys in place of

CC selenocysteine, may be used for monitoring transfection efficiencies
CC or in the study of heterologous promoter function in transient
CC expression assays. Characterisation of the selenocysteine insertion
CC sequence is useful to effect incorporation of selenocysteine into
CC peptides or proteins to study the effects of the presence of
CC selenocysteine on the properties of such proteins.
SQ Sequence 249 AA;
SQ 13 A; 16 R; 14 N; 10 D; 0 B; 4 C; 14 Q; 13 E; 0 Z; 14 G; 7 H;
SQ 12 I; 30 L; 11 K; 7 M; 15 F; 14 P; 13 S; 8 T; 8 W; 6 Y; 19 V;
SQ 1 Others;

Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXLX
|
MGLPQGLWLKRLWLVLEVAHVHV
10 X 20

11. US-08-121-713B-30 (1-7)
R33383 Cytokine mG-CSF.

ID R33383 standard; peptide; 208 AA.
AC R33383;
DT 15-JUL-1993 (first entry)
DE Cytokine mG-CSF.
KW Cytokine; family; leukemia inhibitory factor; LIF; interleukin-6; OSM;
KW granulocyte colony stimulating factor; G-CSF; oncostatin-M; hybrid;
KW IL-6; alpha-helix.
OS Mus musculus.
PN W09305169-A.
PD 18-MAR-1993.
PF 24-AUG-1992; 007112.
PR 30-AUG-1991; US-753178.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI Rose TM, Todaro GJ;
DR WPI; 93-100991/12.
PT Hybrid cytokine(s) contg. four helical regions - derived from
PT LIF, G-CSF, IL-6 or oncostatin-M, useful for treating Kaposi's
PT sarcoma, rheumatoid arthritis, malignancies etc.
PS Disclosure; Fig 1; 41pp; English.
CC The sequences in R33378-85 cytokines derived from the cytokine
CC family members leukemia inhibitory factor (LIF), granulocyte colony
CC stimulating factor (G-CSF), interleukin-6 (IL-6) and oncostatin-M
CC (OSM). These cytokines were used in the production of hybrid
CC cytokines which have unique physiological properties. The hybrid
CC cytokines comprise a first, second, third and fourth alpha-helical
CC region derived from the corresponding region of one of the cytokines
CC given each derived from a factor different from that which at least
CC one additional region of the cytokine is derived.
SQ Sequence 208 AA;
SQ 24 A; 8 R; 2 N; 2 D; 0 B; 5 C; 21 Q; 7 E; 0 Z; 12 G; 5 H;
SQ 5 I; 43 L; 5 K; 5 M; 5 F; 12 P; 21 S; 9 T; 3 W; 4 Y; 10 V;
Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX
|
RMKLMALQLLWQSALWVGREAVPLVT
10 20 X 30

12. US-08-121-713B-30 (1-7) Human granulocyte colony stimulating factor

ID P90051 standard; protein; 207 AA.
AC P90051;
DT 1-NOV-1989 (first entry)
DE Human granulocyte colony stimulating factor
KW Human granulocyte colony stimulating factor; prevents
KW Pseudomonas aeruginosa; Listeria; Candida.
OS Homo sapiens (Human)
PN J01110629-A.
PD 11-MAY-1989.
PF 07-FEB-1986; 501027.
PR 07-FEB-1986; JP-501027.
PA (CHUS) Chugai Pharmaceutical KK.
DR WPI; 89-169399/23.
PT Agent for preventing infection caused eg by
PT Pseudomonas aeruginosa - contg. human granulocyte
PT colony stimulating factors.
PS Disclosure; fig 3; 57pp; Japanese.
CC Human granulocyte colony stimulating factor (h-G-CSF)
CC used as active ingredient for preventing infection by Pseudomonas
CC aeruginosa, Candida and Listeria. h-G-CSF is pref.
CC a neutrophil obtd. from, eg CHU-1 or CHU-2. h-G-CSF
CC has mol. wt. approx. 19 kD, isoelectric point approx.
CC 5.5, 5.8 or 6.1 max. absorption at 280 nm, min. at 250 nm.
CC Part of h-G-CSF (amino acids 1 - 175)
CC can also be used as active agent.

Sequence 207 AA; 4 D; 0 B; 5 C; 21 Q; 10 E; 0 Z; 15 G; 6 H;
SQ 24 A; 5 R; 0 N; 4 I; 39 L; 5 K; 6 M; 6 F; 15 P; 17 S; 9 T; 4 W; 3 Y; 9 V;
SQ 4 I; 39 L; 5 K; 6 M; 6 F; 15 P; 17 S; 9 T; 4 W; 3 Y; 9 V;
Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX
|
PMKLMALQLLWHSALWTVQETPLGP
10 20 X 30

13. US-08-121-713B-30 (1-7) Human G-CSF encoded by pBRG4 insert.

ID P70161 standard; Protein; 207 AA.
AC P70161;
DT 25-APR-1991 (first entry)
DE Human G-CSF encoded by pBRG4 insert.
KW Human granulocyte colony stimulating factor; G-CSF; leukaemia; ss.

OS Homo sapiens.
PN EP-220520-A.
PD 06-MAY-1987.
PF 30-SEP-1986; 113446.
PR 30-SEP-1985; JP-217150.
PR 17-JUL-1986; JP-166710.
PR 17-JUL-1986; JP-166709.
PA (CHUS) Chugai Seiyaku KK.
PI Yamazaki T, Nagata S, Tsuchiya M.
DR WPI; 87-124182/18.
DR N-PSDB; N70224.

PT Polypeptide with human granulocyte colony stimulating factor
PT activity - is obtd. by cultivating transformant formed by
PT recombinant DNA procedures.
PS Disclosure; Fig 3; 73pp; English.
CC The plasmid was isolated from a cDNA prepd. from CHU-2 cells, a
CC human oral cavity tumour cell line, using a 600 bp insert from
CC pHCS-1 (N70221) as a probe. The plasmid was used to prepare
CC recombinant expression plasmids for the prodn. of h G-CSF.
CC See also P70162 and P70163.

SQ Sequence 207 AA;
SQ 24 A; 5 R; 0 N; 4 D; 0 B; 5 C; 20 Q; 11 E; 0 Z; 15 G; 6 H;
SQ 4 I; 39 L; 5 K; 6 M; 6 F; 15 P; 17 S; 9 T; 4 W; 3 Y; 9 V;
SQ 4 I; 39 L; 5 K; 6 M; 6 F; 15 P; 17 S; 9 T; 4 W; 3 Y; 9 V;

Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX
|
PMKLMALQLLWHSALWTVQETPLGP
10 20 X 30

14. US-08-121-713B-30 (1-7) Human G-CSF deduced from genomic DNA.

ID P70163 standard; Protein; 207 AA.
AC P70163;
DT 10-MAR-1993 (revised)
DT 25-APR-1991 (first entry)
DE Human G-CSF deduced from genomic DNA.
KW Human granulocyte colony stimulating factor; G-CSF; leukaemia; ss.
OS Homo sapiens.
PN EP-220520-A.
PD 06-MAY-1987.
PF 30-SEP-1986; 113446.
PR 30-SEP-1985; JP-217150.
PR 17-JUL-1986; JP-166710.
PR 17-JUL-1986; JP-166709.
PA (CHUS) Chugai Seiyaku KK.
PI Yamazaki T, Nagata S, Tsuchiya M.
DR WPI; 87-124182/18.
DR N-PSDB; N70224.

PT Polypeptide with human granulocyte colony stimulating factor
PT activity - is obtd. by cultivating transformant formed by
PT recombinant DNA procedures.
PS Disclosure; Fig 5; 73pp; English.
CC The plasmid was isolated from a human chromosomal DNA library using

CC using a 600 bp insert pHCs-1 isolated from a cDNA library prep.
CC from CHO-2 cells, a human oral cavity tumour cell line.
CC The plasmid was used to prepare recombinant expression plasmids for
CC the prodn. of h G-CSF.
CC See also P70161 and P70162.
SQ Sequence 207 AA;
SQ 24 A; 5 R; 0 N; 4 D; 0 B; 5 C; 20 Q; 11 E; 0 Z; 15 G; 6 H;
SQ 4 I; 39 L; 5 K; 6 M; 6 F; 15 P; 17 S; 9 T; 4 W; 3 Y; 9 V;
Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

```
      X   X
      XWXXXLX
      |
PMKLMALQLLWHSALWTVQEATPLGP
10      20      X      30
```

15. US-08-121-713B-30 (1-7)
P71383 Sequence of human granulocyte colony stimulating f

ID P71383 standard; Protein; 207 AA.
AC P71383;
DT 23-APR-1991 (first entry)
DE Sequence of human granulocyte colony stimulating factor (hGCSF)
DE in pBRG4.
KW Leukopenia therapy; neutropenia; eosinopenia; lymphopenia.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..30
FT Protein 31..208
FT /note= "claimed; N-terminal Met optional"
FT Misc_difference 66..68
FT /note= "optional in claim"
PN EP-217404-A.
PD 08-APR-1987.
PF 03-OCT-1986; 113671.
PR 04-OCT-1985; JP-220450.
PR 02-JUN-1986; JP-125660.
PA (CHUS) CHUGAI SEIYAKO KK.
PI Tamura M, Nomura H, Hattori K, Ono M;
DR WPI; 87-095431/14.
DR N-PSDB; N71319.
PT Leukopenia treating agent esp. for producing fully mature
PT neutrophils - contains human granulocyte colony stimulating
PT factor obtd. by recombinant DNA methods etc.
PS Disclosure; Fig 2; 34pp; English.
CC The hGCSF has the following properties: (i) mol. wt. 19000 +/- 1000
CC (by SDS-PAGE); (ii) isoelectric pt. at 5.5, 5.8 or 6.1 each +/- 0.1;
CC (iii) UV max. absorption at 280nm and min. at 250nm. Prodn. of the
CC gene, vector etc. is described in JP 269455, 269456, 270838 and
CC 270839, each of 1985.
SQ Sequence 207 AA;
SQ 24 A; 5 R; 0 N; 4 D; 0 B; 5 C; 20 Q; 11 E; 0 Z; 15 G; 6 H;
SQ 4 I; 39 L; 5 K; 6 M; 6 F; 15 P; 17 S; 9 T; 4 W; 3 Y; 9 V;
Initial Score = 7 Optimized Score = 7 Significance = 5.71
Residue Identity = 28% Matches = 2 Mismatches = 5

Gaps = 0 Conservative Substitutions = 0
X X
XWXXXLX
|
PMKLMALQLLWHSALWTVQEATPLGP
10 20 X 30
> O <
O | 10 IntelliGenetics
> O <
FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq30pir.res made by on Fri 19 May 95 8:49:28-PDT.

Query sequence being compared: US-08-121-713B-30 (1-7)
Number of sequences searched: 75511
Number of scores above cutoff: 4165

Results of the initial comparison of US-08-121-713B-30 (1-7) with:
Data bank : PIR 43, all entries

```
100000-
-
N
U50000-
M *
B
E
R
O
F10000-
S
E 5000-
Q
D
E
N
C
E
S 1000-
-
500-
-
-
-
-
-
100-
-
50-
-
```



```
#cross-references MUID:89345101
#accession S04844
##molecule_type mRNA
##residues_ 1-2749 ##label FUR
##cross-references EMBL:X15373
##note translation of nucleotide sequence not given
REFERENCE S06796
#authors Furuichi, T.; Yoshikawa, S.; Miyawaki, A.; Wada, K.; Maeda,
N.; Mikoshiba, K.
#journal Nature (1989) 342:32-38
#title Primary structure and functional expression of the inositol
1,4,5-trisphosphate-binding protein P(400).
#cross-references MUID:90044039
#accession S06796
##molecule_type mRNA
##residues_ 1-2749 ##label FUR2
##cross-references EMBL:X15373
##note the nucleotide sequence is not given in this paper
CLASSIFICATION #superfamily inositol-trisphosphate receptor
KEYWORDS glycoprotein; phosphoprotein; receptor; transmembrane protein
FEATURE
890-907 #domain transmembrane #status predicted #label TM1\
1960-1976 #domain transmembrane #status predicted #label TM2\
2276-2294 #domain transmembrane #status predicted #label TM3\
2308-2326 #domain transmembrane #status predicted #label TM4\
2334-2350 #domain transmembrane #status predicted #label TM5\
2352-2372 #domain transmembrane #status predicted #label TM6\
2391-2407 #domain transmembrane #status predicted #label TM7\
2440-2462 #domain transmembrane #status predicted #label TM8\
2570-2589 #domain transmembrane #status predicted #label TM9\
1588,1755 #binding site phosphate (Ser) (covalent) #status
predicted\
2475,2503,2622, #binding site carbohydrate (Asn) (covalent) #status
2710 predicted
SUMMARY #length 2749 #molecular-weight 313195 #checksum 2330
SEQUENCE
Initial Score = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X
X XXXXLX
|
GTLEPHWSGLLWTAMLSIAIVIALPK
2310 X 2320

2. US-08-121-713B-30 (1-7)
A36579 inositol 1,4,5-trisphosphate receptor 1 - rat
ENTRY A36579 #type complete
TITLE inositol 1,4,5-trisphosphate receptor 1 - rat
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change
18-Jun-1993
ACCESSIONS A36579
REFERENCE A36579
#authors Mignery, G.A.; Newton, C.L.; Archer III, B.T.; Suedhof, T.C.
```

```
#journal J. Biol. Chem. (1990) 265:12679-12685
#title Structure and expression of the rat inositol 1,4,
5-trisphosphate receptor.
#cross-references MUID:90324264
#accession A36579
##status preliminary
##molecule_type mRNA
##residues_ 1-2749 ##label MIG
##cross-references GB:J05510
CLASSIFICATION #superfamily inositol-trisphosphate receptor
SUMMARY #length 2749 #molecular-weight 313134 #checksum 2229
SEQUENCE
Initial Score = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X
X XXXXLX
|
GTLEPHWSGLLWTAMLSIAIVIALPK
2310 X 2320

3. US-08-121-713B-30 (1-7)
B36579 inositol 1,4,5-trisphosphate receptor 2 - rat
ENTRY B36579 #type complete
TITLE inositol 1,4,5-trisphosphate receptor 2 - rat
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change
18-Jun-1993
ACCESSIONS B36579
REFERENCE A36579
#authors Mignery, G.A.; Newton, C.L.; Archer III, B.T.; Suedhof, T.C.
#journal J. Biol. Chem. (1990) 265:12679-12685
#title Structure and expression of the rat inositol 1,4,
5-trisphosphate receptor.
#cross-references MUID:90324264
#accession B36579
##status preliminary
##molecule_type mRNA
##residues_ 1-2734 ##label MIG
##cross-references GB:J05510
CLASSIFICATION #superfamily inositol-trisphosphate receptor
SUMMARY #length 2734 #molecular-weight 311368 #checksum 6817
SEQUENCE
Initial Score = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X
X XXXXLX
|
GTLEPHWSGLLWTAMLSIAIVIALPK
2290 2300 X 2310

4. US-08-121-713B-30 (1-7)
```



```

A04743      IP3 receptor, XIP3R - African clawed frog

ENTRY       A40743      #type complete
TITLE       IP3 receptor, XIP3R - African clawed frog
ORGANISM    #formal name Xenopus laevis #common name African clawed frog
DATE        03-Mar-1994 #sequence_revision 18-Nov-1994 #text_change
              18-Nov-1994

ACCESSIONS  A40743
REFERENCE    A40743
AUTHORS      Kume, S.; Muto, A.; Aruga, J.; Nakagawa, T.; Michikawa, T.;
              Furuichi, T.; Nakade, S.; Okano, H.; Mikoshiba, K.
JOURNAL      Cell (1993) 73:555-570
TITLE       The Xenopus IP3 receptor: structure, function, and
              localization in oocytes and eggs.
CROSS-REFERENCES MUID:93258819
CONTENTS     oocytes and eggs
ACCESSION    A40743
STATUS       #status preliminary
MOLECULE     #molecule_type mRNA
RESIDUES     1-2693 #label KDM
CROSS-REFERENCES NCBIP:131713
              sequence extracted from NCBI backbone
              sequence not compared to nucleotide translation
              #note
              #note
SUMMARY      #length 2693 #molecular-weight 306670 #checksum 918
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

      X      X
      XWXXLX
      |
      2260 X 2270
      GTTDSRLSGLLWTAMLVSLAIVIVLPK

5. US-08-121-7113B-30 (1-7)
A46692      DNA polymerase epsilon catalytic subunit - human

ENTRY       A46692      #type complete
TITLE       DNA polymerase epsilon catalytic subunit - human
ORGANISM    #formal name Homo sapiens #common name man
DATE        21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
              18-Nov-1994

ACCESSIONS  A46692
REFERENCE    A46692
AUTHORS      Kesti, T.; Frantti, H.; Syvaoja, J.E.
              J. Biol. Chem. (1993) 268:10238-10245
JOURNAL      Molecular cloning of the cDNA for the catalytic subunit of
              human DNA polymerase epsilon.
CONTENTS     HeLa cells
ACCESSION    A46692
STATUS       #status preliminary
MOLECULE     #molecule_type nucleic acid; protein
RESIDUES     1-2257 #label KES
CROSS-REFERENCES NCBIN:131270; NCBIP:131271
              sequence inconsistent with the nucleotide translation
              sequence extracted from NCBI backbone
              #note
              #note
SUMMARY      #length 2257 #molecular-weight 257967 #checksum 1781
SEQUENCE

```

```

SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

      X      X
      XXXXX
      |
FHSLTISFRCWEFLMMDPSPNYGIIK
1880      1990

6. US-08-121-713B-30 (1-7)
   RXPPLC   RNA-directed RNA polymerase (EC 2.7.7.48) - lympho

ENTRY      #type complete
TITLE      RNA-directed RNA polymerase (EC 2.7.7.48) - lymphocytic
           choriomeningitis virus (strain Armstrong 53b)
ORGANISM   #formal name lymphocytic choriomeningitis virus
DATE       31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
           17-Feb-1994
ACCESSIONS A30181
REFERENCE   A30181
            Salvato, M.; Shimomaye, E.; Oldstone, M.B.A.
            Virology (1989) 169:377-384
            The primary structure of the lymphocytic choriomeningitis
            virus L gene encodes a putative RNA polymerase.
            #cross-references MUID:89204909
ACCESSION  A30181
            #molecule_type genomic RNA
            #residues      1-2210 #label SAL

GENETICS
            #map_position segment L
            CLASSIFICATION #superfamily arenavirus RNA-directed RNA polymerase
            KEYWORDS        nucleotidyltransferase
            SUMMARY         #length 2210 #molecular-weight 254529 #checksum 7158
            SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

```

```

NPKAFKSIKOLWYMLNYTKGVLEFSI
1930      1940      X 1950

7. US-08-121-713B-30 (1-7)
   BWSABE      bImE protein - Emericella nidulans

ENTRY      BWSABE      #type complete
TITLE      bImE protein - Emericella nidulans
ORGANISM   #formal name Emericella nidulans, Aspergillus nidulans
DATE       31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
           09-Sep-1994
ACCESSIONS A37879
REFERENCE   A37879

```

#authors Engle, D.B.; Osmani, S.A.; Osmani, A.H.; Rosborough, S.;
Xiang, X.; Morris, N.R.
#journal J. Biol. Chem. (1990) 263:16132-16137
#title A negative regulator of mitosis in Aspergillus is a putative
membrane-spanning protein.
#cross-references MUID:90375468
#accession A37879

##molecule_type mRNA
##residues 1-2073 ##label ENG
##cross-references GB:J05607
##note in addition to three predicted transmembrane domains,
there are several potential N-glycosylation sites,
five possible sites for phosphorylation by
cAMP-dependent protein kinase and one for casein
kinase, and one sequence that resembles a nuclear
localization signal

COMMENT This protein is part of a regulatory pathway that includes the nima
protein kinase. It is required to prevent premature entry into
mitosis. Mutations to this protein both cause cells to enter
mitosis and prevent them from leaving mitosis.

GENETICS

#gene bImE
#superfamily bImE protein
KEYWORDS cell cycle control; mitosis; transmembrane protein
FEATURE

1623-1643 #domain transmembrane #status predicted #label TM1\
1685-1703 #domain transmembrane #status predicted #label TM2\
1746-1764 #domain transmembrane #status predicted #label TM3
SUMMARY #length 2073 #molecular-weight 229177 #checksum 7408
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX
|
SSSGSYVDTSWESMLFQESGAGVVA
850 X 860 870

8. US-08-121-713B-30 (1-7)
A40262

210K nuclear DNA-binding protein CCG1 - human

ENTRY A40262 #type complete
TITLE 210K nuclear DNA-binding protein CCG1 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change
02-Dec-1994

ACCESSIONS A40262; S03005; S00830
REFERENCE A40262
#authors Sekiguchi, T.; Nohiro, Y.; Nakamura, Y.; Hisamoto, N.;
Nishimoto, T.

#journal Mol. Cell. Biol. (1991) 11:3317-3325
#title The human CCG1 gene, essential for progression of the G-1
phase, encodes a 210-kilodalton nuclear DNA-binding
protein.

#cross-references MUID:91246200
#accession A40262

##molecule_type mRNA
##residues 1-1872 ##label SEK
##cross-references GB:D90359
##note nucleotide sequence not complete in this paper
REFERENCE S03005

#authors Sekiguchi, T.; Miyata, T.; Nishimoto, T.
#submission submitted to the EMBL Data Library, February 1988
#accession S03005
##molecule_type mRNA
##residues 'MYR', 60-1583, 'DN', 1586, 'CSSKANDIVCLIQCSSQI', 1606,
'ELRF', ##label SE2

##cross-references EMBL:X07024

##note this sequence has been revised in reference A40262

REFERENCE S00830
#authors Sekiguchi, T.; Miyata, T.; Nishimoto, T.
#journal EMBO J. (1988) 7:1683-1687
#title Molecular cloning of the cDNA of human X chromosomal gene
(CCG1) which complements the temperature-sensitive G1
mutants, tsBN462 and ts13, of the BHK cell line.

#cross-references MUID:89005056

#accession S00830

##molecule_type DNA

##residues 1351-1358; 1365-1583, 'DN', 1586, 'CSSKANDIVCLIQCSSQI',
1606, 'ELRF', ##label SE3

##cross-references EMBL:X07024

GENETICS

#gene GDB:CCG1
#map_position Xq13.1
CLASSIFICATION #superfamily bromodomain homology
FEATURE

1405-1460 #domain bromodomain homology #label BR01\
1528-1583 #domain bromodomain homology #label BR02
SUMMARY #length 1872 #molecular-weight 212675 #checksum 1180
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX
|

VAEWRYGPARLWYDMLGVPEGSGFDY
340 350 X 360

9. US-08-121-713B-30 (1-7)

S23810 collagen alpha 1(XVI) chain precursor - human

ENTRY S23810 #type complete
TITLE collagen alpha 1(XVI) chain precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change
28-Oct-1994

ACCESSIONS S23810; S08012
REFERENCE S23810

#authors Pan, T.-C.; Zhang, R.-Z.; Mattei, M.G.; Timpl, R.; Chu, M.L.
#journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:6565-6569
#title Cloning and chromosomal location of human alpha(XVI)
collagen.

```
#cross-references MWID:92335339
#accession S23810
##molecule_type mRNA
##residues 1-1603 ##label PAN
##cross-references EMBL:M92642
REFERENCE
#authors Kimura, S.
#submission submitted to the EMBL Data Library, April 1989
#description Partial nucleotide and amino acid sequence of a collagen-like
protein from human placenta.
#accession S08012
##molecule_type mRNA
##residues 403-419, 'GR', 421-536, 'P', 538-846, 'VM' ##label KIM
##cross-references EMBL:X14963
CLASSIFICATION #superfamily collagen alpha chain
KEYWORDS cell binding; extracellular matrix; glycoprotein;
hydroxylysine; hydroxyproline
FEATURE
1-21 #domain signal sequence #status predicted #label SIG\
22-1603 #product collagen alpha 1(XVI) chain #status predicted
#label MAT\
334-360 #domain collagenous COL10 #label COL10\
375-505 #domain collagenous COL9 #label COL9\
521-554 #domain collagenous COL8 #label COL8\
539-541 #region cell attachment (R-G-D) motif\
572-630 #domain collagenous COL7 #label COL7\
652-722 #domain collagenous COL6 #label COL6\
738-875 #domain collagenous COL5 #label COL5\
887-938 #domain collagenous COL4 #label COL4\
973-987 #domain collagenous COL3 #label COL3\
1005-1007 #region cell attachment (R-G-D) motif\
1011-1432 #domain collagenous COL2 #label COL2\
1226-1228 #region cell attachment (R-G-D) motif\
1472-1577 #domain collagenous COL1 #label COL1\
47,527 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 1603 #molecular-weight 157693 #checksum 1714
SEQUENCE
Initial Score = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X
XWXXXIX
|
MWVSWAPGLWLLGLWATFGHGANTG
10 X 20
#polycystic kidney disease protein 1 - human (fragm

10. US-08-121-713B-30 (1-7)
A44604 polycystic kidney disease protein 1 - human (fragment)
ENTRY #type fragment
TITLE polycystic kidney disease protein 1 - human (fragment)
ORGANISM #formal name Homo sapiens #common name man
DATE 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change
17-Jul-1994
ACCESSIONS A44604
REFERENCE A44604
```

```
#authors Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.;
MacCarthy, A.B.; Sloane-Stanley, J.; Buckle, V.J.; Kearney,
L.; Higgs, D.R.; Ratcliffe, P.J.; Harris, P.C.; Roelfsema,
J.H.; Spruit, L.; Saris, J.J.; Dauwerse, H.G.; Peters,
D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter, P.T.;
Maneshwar, M.M.; Cordeliro, I.; Santos, H.; Cabral, P.;
Sampson, J.R.; Janssen, B.; Hesselting-Janssen, A.L.W.; van
den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout,
D.; Hallley, D.J.J.
#journal Cell (1994) 77:881-894
#title The polycystic kidney disease 1 gene encodes a 14 kb
transcript and lies within a duplicated region on
chromosome 16.
#accession A44604
##status preliminary
##molecule_type mRNA
##residues 1-1594 ##label WAR
##cross-references GB:L33243
SUMMARY #length 1594 #checksum 7489
SEQUENCE
Initial Score = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0
X X
XWXXXIX
|
WHLSPLLCVGLWALRLWGLALRLGAVIL
1300 1310

11. US-08-121-713B-30 (1-7)
I48216 neuexin III-alpha membrane-bound type 1 precursor
ENTRY #type complete
TITLE neuexin III-alpha membrane-bound type 1 precursor - rat
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 26-May-1994 #sequence_revision 26-May-1994 #text_change
26-May-1994
ACCESSIONS I48216
REFERENCE A48216
#authors Ushkaryov, Y.A.; Suedhof, T.C.
#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:6410-6414
#title Neuexin IIIalpha: extensive alternative splicing generates
membrane-bound and soluble forms.
#accession I48216
##status preliminary
##molecule_type mRNA
##residues 1-1578 ##label USH
##cross-references GB:L14851
GENETICS 1372/1
KEYWORDS #introns
alternative splicing; brain; cell surface component;
receptor; repeat; transmembrane protein
FEATURE #domain signal sequence #status predicted #label SIG
1-27 #length 1578 #molecular-weight 174212 #checksum 5082
SUMMARY
SEQUENCE
```

```

Initial Score      = 7  Optimized Score = 7  Significance = 3.94
Residue Identity = 28%  Matches          = 5
Gaps              = 0    Conservative Substitutions = 0

```

```

      X      X
      XXXXX
      |
NRAGILPTELTAMLNLYGYVGCIKDL
600 X      610

```

12. US-08-121-713B-30 (1-7)
neurexin III-alpha membrane-bound type 2 precursor

```

ENTRY      A48218      #type complete
TITLE      neurexin III-alpha membrane-bound type 2 precursor - rat
ORGANISM   #formal name Rattus norvegicus #common name Norway rat
DATE       26-May-1994 #sequence_revision 26-May-1994 #text_change
           26-May-1994
ACCESSIONS A48218
REFERENCE   A48216
            Ushkaryov, Y.A.; Suedhof, T.C.
            Proc. Natl. Acad. Sci. U.S.A. (1993) 90:6410-6414
            Neurexin IIIalpha: extensive alternative splicing generates
            membrane-bound and soluble forms.
            #accession A48218
            #status preliminary
            #molecule_type mRNA
            #residues 1-1575 #label USH
            #cross-references GB:114851

```

12. US-08-121-713B-30 (1-7)
A48218 neurexin III-alpha membrane-bound type 2 precursor

```
ENTRY      A48218          #type complete
TITLE      neuroxin iii-alpha membrane-bound type 2 precursor - rat
ORGANISM   #formal name Rattus norvegicus #common name Norway rat
DATE       26-May-1994   #sequence_revision 26-May-1994 #text_change
                        26-May-1994

ACCESSIONS A48218
REFERENCE   A48216
AUTHORS    Ushkaryov, Y.A.; Suedhof, T.C.
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. (1993) 90:6410-6414
TITLE       Neuroxin IIIalpha: extensive alternative splicing generates
            membrane-bound and soluble forms.

ACCESSION  A48218
STATUS      preliminary
MOLECULE    type mRNA
RESIDUES    1-1575  #label USH
CROSS-REFS  #cross-references GB:L14851
```

KEYWORDS
#introns
1369/1
alternative splicing; brain; cell surface component;
receptor; repeat; transmembrane protein

```

FEATURE
1-27      #domain signal sequence #status predicted #label SIG
          #length 1575 #molecular-weight 173911 #checksum 6342
SUMMARY
SEQUENCE

```

```
Initial Score      = 7 Optimized Score = 7 Significance = 3.94
Residue Identity = 28% Matches      = 2 Mismatches = 5
Gaps              = 0 Conservative Substitutions = 0
```

X X
 XWXXXLX
 |
 NRAGLIPLTELWTAMLYN
 600 X 61

13. US-08-121-713B-30 (1-7)
neurexin III-alpha membrane-bound type 3 precursor
B48218

ENTRY	B48218	#type complete
TITLE	neurexin III-alpha membrane-bound type 3 precursor	- rat
ORGANISM	Neuraln	name Rattus norvegicus #common name Norway rat
DATE	26-May-1994	#sequence_revision 26-May-1994 #text_change 26-May-1994
ACCESSIONS	B48218	
REFERENCE	A48216	

#authors Ushkaryov, Y.A.; Suedhof, T.C.
#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:6410-6414
#title Neurexin IIIalpha: a distinct alternative splicing generates membrane-bound and soluble forms.
#accession B48218
#status preliminary
#molecule_type mRNA
#residues 1-1471
#label USH
#Cross-references GB:L14851

GENETICS #introns 1372/1

KEYWORDS
alternative splicing; brain; cell surface component; receptor; repeat; transmembrane protein

```

#domain signal sequence #status predicted #label SIG
1-27
SUMMARY
SEQUENCE
#length 1471 #molecular-weight 162434 #checksum 183

```

Initial Score	=	7	Optimized Score	=	7	Significance	=	3.94
Residue Identity	=	28%	Matches	=	28	Mismatches	=	5
Gaps	=	0	Conservative Substitutions	=	0		=	0

NRAGLILPTELWTAMLN
600 · X 61

NRAGLILPTELWTAMLNYYGVGCIRDL
600 · X 610

4. US-08-121-713B-30 (1-7)
C48218 neurexin III-alpha membrane-bound type 4 precursor

ENTRY	C48218	#type complete
TITLE	neurexin III-alpha membrane-bound type 4 precursor - rat	
ORGANISM	#formal name Rattus norvegicus #common name Norway rat	
DATE	26-May-1994	#sequence_revision 26-May-1994 #text_change 26-May-1994

ACCESSIONS	REFERENCE	#authors	#journal	#title
C48218	Ushkaryov, Y.A.; Suedhof, T.C.			
A48216	Proc. Natl. Acad. Sci. U.S.A. (1993) 90:6410-6414			
	Neurexin IIIalpha: extensive alternative splicing generates			
	membrane-bound and soluble forms.			

```

#accession      C48218      preliminary
##status
##molecule_type mRNA
##residues      1-1468    ##label USH
##cross-references GB:L14851

```

GENETICS
#introns
KEYWORDS
1369/1
alternative splicing; brain; cell surface component;
receptor; repeat; transmembrane protein

```

#domain signal sequence #status predicted #label SIG
1-27
SUMMARY
#length 1468 #molecular-weight 162134 #checksum 408

```

Initial Score	=	7	Optimized Score	=	7	Significance	=	3.94
Residue Identity	=	28%	Matches	=	2	Mismatches	=	5
Gaps	=	0	Conservative Substitutions	=	0		=	0

X
X

X
X
X
X
L
X

NRAGLILPTELWTAMLN¹GYVGCIRDL
600 X 610

```

15. US-08-121-713B-30 (1-7)
   A45665 adult-specific brush border esterase/phospholipase

ENTRY      A45665 #type complete
TITLE      adult-specific brush border esterase/phospholipase (EC
           3.---) precursor - rabbit
ORGANISM   #formal name Oryctolagus cuniculus #common name domestic
           rabbit
DATE       03-May-1994 #sequence_revision 03-May-1994 #text_change
           03-May-1994
ACCESSIONS A45665
REFERENCE   A45665
#authors    Boll, W.; Schmid-Chanda, T.; Semenza, G.; Mantei, N.
#journal     J. Biol. Chem. (1993) 268:12901-12911
#title       Messenger RNAs expressed in intestine of adult but not baby
           rabbits. Isolation of cognate cDNAs and characterization of
           a novel brush border protein with esterase and
           phospholipase activity.
#accession  A45665
#status      preliminary
#molecule_type mRNA
#residues    length 1-1458 ##label BOL
#cross-references GB:212841
#length 1458 #molecule-weight 161343 #checksum 2780

SUMMARY
SEQUENCE

```

Initial Score	=	7	Optimized Score	=	7	Significance	=	3.94
Residue Identity	=	28%	Matches	=	2	Mismatches	=	5
Gaps	=	0	Conservative Substitutions	=	0			0

X
X
XTXXXXMX

SKAHAAASALWNNMLEPVGQKTHND
660 670 X 680

IntelliGenetics

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq30spt.res made by on Fri 19 May 95 9:00:53-PDT.

Query sequence being compared: US-08-121-713B-30 (1-7)
 Number of sequences searched: 43470
 Number of scores above cutoff: 4001

Results of the initial comparison of US-08-121-713B-30 (1-7) with:
Data bank : Swiss-Prot31, all entries

Category	Number of Studies
N	100000
U	50000
M	10000
B	5000
E	1000
R	500
O	100
F	50
S	10
E	5
Q	1
U	1
E	1
N	1
C	1
E	1
S	1

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20

Gap penalty 1.00 Window size 6
Gap size penalty 0.05
Cutoff score 0
Randomization group 0
Initial scores to save 45 Alignments to save 15
Optimized scores to save 0 Display context 10
SEARCH STATISTICS
Scores: Mean Median Standard Deviation
1 4 1.50
Times: CPU Total Elapsed
00:00:43.06 00:00:44.00
Number of residues: 15335248
Number of sequences searched: 43470
Number of scores above cutoff: 4001
Cut-off raised to 1.
Cut-off raised to 3.
Cut-off raised to 4.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

71 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
1. IP3R_MOUSE	INOSITOL 1,4,5-TRISPHOSPHATE-	2749	7	7	3.99 0
2. IP3R_RAT	INOSITOL 1,4,5-TRISPHOSPHATE-	2749	7	7	3.99 0
3. DPGE_HUMAN	DNA POLYMERASE EPSILON, CATAL	2257	7	7	3.99 0
4. RPOE_LYCA	RNA POLYMERASE (EC 2.7.7.48).	2210	7	7	3.99 0
5. BIME_EMENI	NEGATIVE REGULATOR OF MITOSIS	2073	7	7	3.99 0
6. CCG1_HUMAN	TRANSCRIPTION INITIATION FACT	1872	7	7	3.99 0
7. CALF_HUMAN	COLLAGEN ALPHA 1(XVI) CHAIN P	1603	7	7	3.99 0
8. PHLX_RABIT	PHOSPHOLIPASE ADRA-B PRECURS	1458	7	7	3.99 0
9. EBAL_PLAFC	ERYTHROCYTE-BINDING ANTIGEN E	1426	7	7	3.99 0
10. KRE5_YEAST	KILLER TOXIN-RESISTANCE PROTE	1365	7	7	3.99 0
11. YB95_YEAST	HYPOTHETICAL 132.7 KD HELICAS	1143	7	7	3.99 0
12. SP23_YEAST	SPT23 PROTEIN.	1082	7	7	3.99 0
13. PPS3_BACSD	PEPTIDE SYNTHETASE 3 (FRAGMEN	859	7	7	3.99 0
14. YLH5_CAEEL	POTATIVE ABC TRANSPORTER C48B	856	7	7	3.99 0
15. SYFB_BACSU	PHENYLALANYL-TRNA SYNTHETASE	804	7	7	3.99 0
16. APRF_HUMAN	ACUTE-PHASE RESPONSE FACTOR.	770	7	7	3.99 0
17. AID_HUMAN	ADRENOLUKODYSTROPHY PROTEIN	745	7	7	3.99 0
18. ACPE_RAT	ACYLAMINO-ACID-RELEASING ENZY	732	7	7	3.99 0
19. DCP_ECOLI	DIPETIDYL CARBOXYPEPTIDASE (680	7	7	3.99 0
20. DCP_SALTY	DIPEPTIDYL CARBOXYPEPTIDASE	680	7	7	3.99 0
21. APG_ARATH	ANTER-SPECIFIC PROLINE-RICH P	534	7	7	3.99 0
22. YUCC_ECOLI	HYPOTHETICAL 60.8 KD PROTEIN	528	7	7	3.99 0
23. COX1_APTME	CYTOCHROME C OXIDASE POLYPEPT	521	7	7	3.99 0
24. NQ2M_BETVU	NADH-UBIQUINONE OXIDOREDUCTAS	515	7	7	3.99 0
25. NUOM_ECOLI	NADH-DEHYDROGENASE 1 CHAIN M	509	7	7	3.99 0
26. PC06_MOUSE	PROTEIN PC0-6 (FRAGMENT).	500	7	7	3.99 0
27. YPUM_RHOCA	HYPOTHETICAL 50.4 KD PROTEIN	477	7	7	3.99 0

28. PRI1_MOUSE DNA PRIMASE 49 KD SUBUNIT (EC 417 7 3.99 0
29. CVAA_ECOLI COLICIN V SECRETION PROTEIN C 413 7 3.99 0
30. YBL2_SFV3L BEL-2 PROTEIN. 388 7 3.99 0
31. YGIC_ECOLI HYPOTHETICAL 45.0 KD PROTEIN 386 7 3.99 0
32. CAR2_DICDI CYCLIC AMP RECEPTOR 2. 375 7 3.99 0
33. YEJP_ECOLI HYPOTHETICAL 39.1 KD PROTEIN 350 7 3.99 0
34. YIFC_ECOLI HYPOTHETICAL 39.6 KD PROTEIN 349 7 3.99 0
35. SUR4_YEAST SUR4 PROTEIN (SRE1 PROTEIN). 345 7 3.99 0
36. YIBD_ECOLI HYPOTHETICAL 40.5 KD PROTEIN 344 7 3.99 0
37. TRAU_MOUSE TRAU PROTEIN PRECURSOR. 330 7 3.99 0
38. UPAR_MOUSE UROKINASE PLASMINOGEN ACTIVAT 327 7 3.99 0
39. SYEH_HUMAN SYNAPTOPHYSIN (MAJOR SYNAPTIC 313 7 3.99 0
40. SYEH_BOVIN SYNAPTOPHYSIN (MAJOR SYNAPTIC 307 7 3.99 0
41. SYEH_RAT SYNAPTOPHYSIN (MAJOR SYNAPTIC 284 7 3.99 0
42. YHBJ_ECOLI HYPOTHETICAL 32.5 KD PROTEIN 284 7 3.99 0
43. YF11_STRPA HYPOTHETICAL 29.9 KD PROTEIN 280 7 3.99 0
44. YPU2_RHOCA HYPOTHETICAL 30.4 KD PROTEIN 274 7 3.99 0
45. NODJ_RHILT NODULATION PROTEIN J. 262 7 3.99 0

1. US-08-121-713B-30 (1-7)

IP3R_MOUSE INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN TYPE

ID IP3R_MOUSE STANDARD; PRT; 2749 AA.
AC P11881;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN TYPE 1 RECEPTOR
DE (INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN P(400)) (TYPE 1 INSP3
DE RECEPTOR).
GN INSP3R.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CERBELLA PURKINJE NEURONS;
RM 90044039
RA FURUICHI T., YOSHIKAWA S., MIYAWAKI A., WADA K., MAEDA N.,
RA MIKOSHIBA K.;
RL NATURE 342:32-38 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=CEREBELLUM;
RM 89345101
RA FURUICHI T., YOSHIKAWA S., MIKOSHIBA K.;
RL NUCLEIC ACIDS RES. 17:5385-5386 (1989).
RN [3]
RP ALTERNATIVE SPLICING.
RM 91296797
RA NAKAGAWA T., OKANO H., FURUICHI T., ARUGA J., MIKOSHIBA K.;
RL PROC. NATL. ACAD. SCI. U.S.A. 88:6244-6248 (1991).
CC -!- FUNCTION: RECEPTOR FOR INOSITOL 1,4,5-TRISPHOSPHATE, A SECOND
CC MESSENGER THAT MEDIATES THE RELEASE OF INTRACELLULAR CALCIUM.
CC THE RECEPTOR CONTAINS A CALCIUM CHANNEL IN ITS C-TERMINAL
CC EXTREMITY. ITS LARGE N-TERMINAL CYTOPLASMIC REGION HAS THE LIGAND-
CC BINDING SITE IN THE N-TERMINUS AND MODULATORY SITES IN THE MIDDLE
CC PORTION IMMEDIATELY UPSTREAM OF THE CHANNEL REGION.
CC -!- SUBUNIT: HOMOTETRAMER.

CC -1- PTM: PHOSPHORYLATED BY CYCLIC-AMP KINASE. PHOSPHORYLATION PREVENTS
 CC THE LIGAND-INDUCED OPENING OF THE CALCIUM CHANNELS.
 CC -1- CALCIUM APPEARS TO INHIBIT LIGAND BINDING TO THE RECEPTOR, MOST
 CC PROBABLY BY INTERACTING WITH A DISTINCT CALCIUM-BINDING PROTEIN
 CC WHICH THEN INHIBITS THE RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- ALTERNATIVE PRODUCTS: ADDITIONAL SUBTYPES OF INSP3R ARISE BY
 CC ALTERNATIVE SPLICING OF THE SAME GENE.
 CC -1- SIMILARITY: TO RYANODINE RECEPTOR.
 DR EMBL; X15373; MPM400.
 DR PIR; S04844; ACM517.
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION;
 KW ENDOPLASMIC RETICULUM; IONIC CHANNEL; ION TRANSPORT; CALCIUM CHANNEL;
 KW ALTERNATIVE SPLICING.
 FT DOMAIN 1 2273 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 2274 2294 M1 (POTENTIAL).
 FT TRANSMEM 2308 2326 M2 (POTENTIAL).
 FT TRANSMEM 2334 2356 M3 (POTENTIAL).
 FT TRANSMEM 2365 2387 M4 (POTENTIAL).
 FT TRANSMEM 2391 2407 M5 (POTENTIAL).
 FT TRANSMEM 2440 2462 M6 (POTENTIAL).
 FT TRANSMEM 2530 2549 M7 (POTENTIAL).
 FT TRANSMEM 2570 2589 M8 (POTENTIAL).
 FT DOMAIN 2590 2749 CYTOPLASMIC (POTENTIAL).
 FT VARSPIC 318 332 MISSING (IN INSP3R SI-).
 FT VARSPIC 1692 1731 MISSING (IN INSP3R SIIBC-).
 FT VARSPIC 1715 1731 MISSING (IN INSP3R SIIBC-).
 FT VARSPIC 1715 1731 MISSING (IN INSP3R SIIBC-).
 FT MOD RES 1588 1598 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
 FT MOD RES 1755 1755 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
 SQ SEQUENCE 2749 AA; 313193 MW; 22837568 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X X
 XWXXXLX
 GTLEPHWSGLLWTAMLSLAIVIALPK
 2310 X 2320

2. US-08-121-713B-30 (1-7)
 IP3R_RAT INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN TYPE

ID IP3R_RAT STANDARD; PRT; 2749 AA.
 AC P29994;
 DT 01-APR-1993 (REL. 25, CREATED)
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN TYPE 1 RECEPTOR
 DE (INOSITOL 1,4,5-TRISPHOSPHATE-BINDING PROTEIN P(400)) (TYPE 1 INSP3
 DE RECEPTOR).
 GN INSP3R.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]

RP SEQUENCE FROM N.A.
 RM 90324264
 RA MIGNERY G.A., NEWTON C.L., ARCHER B.T. III, SUEHOF T.C.;
 RL J. BIOL. CHEM. 265:12679-12685(1990).
 CC -1- FUNCTION: RECEPTOR FOR INOSITOL 1,4,5-TRISPHOSPHATE, A SECOND
 CC MESSENGER THAT MEDIATES THE RELEASE OF INTRACELLULAR CALCIUM.
 CC THE RECEPTOR CONTAINS A CALCIUM CHANNEL IN ITS C-TERMINAL
 CC EXTREMITY. ITS LARGE N-TERMINAL CYTOPLASMIC REGION HAS THE LIGAND-
 CC BINDING SITE IN THE N-TERMINUS AND MODULATORY SITES IN THE MIDDLE
 CC PORTION IMMEDIATELY UPSTREAM OF THE CHANNEL REGION.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- PTM: PHOSPHORYLATED BY CYCLIC-AMP KINASE. PHOSPHORYLATION PREVENTS
 CC THE LIGAND-INDUCED OPENING OF THE CALCIUM CHANNELS.
 CC -1- CALCIUM APPEARS TO INHIBIT LIGAND BINDING TO THE RECEPTOR, MOST
 CC PROBABLY BY INTERACTING WITH A DISTINCT CALCIUM-BINDING PROTEIN
 CC WHICH THEN INHIBITS THE RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED.
 CC -1- SIMILARITY: TO RYANODINE RECEPTOR.
 DR EMBL; J05510; RRI145TR.
 DR PIR; A36579; A36579.
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION;
 KW ENDOPLASMIC RETICULUM; IONIC CHANNEL; ION TRANSPORT; CALCIUM CHANNEL;
 KW ALTERNATIVE SPLICING.
 FT DOMAIN 1 2273 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 2274 2294 M1 (POTENTIAL).
 FT TRANSMEM 2308 2326 M2 (POTENTIAL).
 FT TRANSMEM 2334 2356 M3 (POTENTIAL).
 FT TRANSMEM 2365 2387 M4 (POTENTIAL).
 FT TRANSMEM 2391 2407 M5 (POTENTIAL).
 FT TRANSMEM 2440 2462 M6 (POTENTIAL).
 FT TRANSMEM 2530 2549 M7 (POTENTIAL).
 FT TRANSMEM 2570 2589 M8 (POTENTIAL).
 FT DOMAIN 2590 2749 CYTOPLASMIC (POTENTIAL).
 FT MOD RES 1589 1589 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
 FT MOD RES 1755 1755 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
 FT VARSPIC 322 336 MISSING (IN VARIANT).
 FT VARIANT 1372 1372 MISSING (IN ALL, BUT P17 CLONES).
 SQ SEQUENCE 2749 AA; 313132 MW; 22831603 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X X
 XWXXXLX
 GTLEPHWSGLLWTAMLSLAIVIALPK
 2310 X 2320

3. US-08-121-713B-30 (1-7)
 DPOE_HUMAN DNA POLYMERASE EPSILON, CATALYTIC SUBUNIT A (EC 2.

ID DPOE_HUMAN STANDARD; PRT; 2257 AA.
 AC Q07864;
 DT 01-OCT-1994 (REL. 30, CREATED)
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)

DE DNA POLYMERASE EPSILON, CATALYTIC SUBUNIT A (EC 2.7.7.7) (DNA
DE POLYMERASE II SUBUNIT A).
GN POLE.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUATHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 48-51; 874-884 AND 1336-1342.
RM 93252906
RA KESTI T., FRANTTI H., SYVAOJA J.E.;
RL J. BIOL. CHEM. 268:10238-10245(1993).
CC -1- FUNCTION: PARTICIPATES IN DNA REPAIR AND IN CHROMOSOMAL DNA
CC REPLICATION.
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + DNA(N).
CC -1- SUBUNIT: CONSISTS OF TWO SUBUNITS (258 KD AND 55 KD).
CC -1- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
CC FOR COMPLEXING SUBUNITS B AND C. THE C-TERMINUS MAY ALSO REGULATE
CC THE CATALYTIC ACTIVITIES OF THE ENZYME.
CC -1- IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES: ALPHA, BETA, GAMMA,
CC DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR DIFFERENT REACTIONS
CC OF DNA SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO FAMILY B OF DNA POLYMERASES. HIGH
CC SIMILARITY WITH YEAST DNA POLYMERASE II.
CC EMBL; L09561; HSDNPOLCS.
DR PIR; A46692; A46692.
DR MIM; 174762; 11TH EDITION.
KW DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION; DNA REPAIR; DNA-BINDING;
KW ZINC-FINGER.
FT ZN FING 2129 2209 POTENTIAL.
SQ SEQUENCE 2257 AA; 257967 MW; 19650612 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXLX
|
FHSLTISFRCWEFLWMDPSNYGGIK
1880 1890

4. US-08-121-713B-30 (1-7)
RRPO_LYCVA RNA POLYMERASE (EC 2.7.7.48).

ID RRPO_LYCVA STANDARD; PRT; 2210 AA.
AC P14240;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE (EC 2.7.7.48).
GN L.
OS LYMPHOCTIC CHORIOMENINGITIS VIRUS (STRAIN ARMSTRONG).
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; NEGATIVE-STRAND; ARENAVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RM 89204909
RA SALVATO M.S., SHIMOMAYE E.M., OLDSTONE M.B.A.;

RL VIROLOGY 169:377-384(1989).
RN [2]
RP SEQUENCE OF 161-387; 424-619 AND 1646-1906 FROM N.A.
RM 88072084
RA SINGH M.K., FULLER-PACE F.V., BUCHMEIER M.J., SOUTHERN P.J.;
RL VIROLOGY 161:448-456(1987).
DR EMBL; J04331; LCVLPY.
DR EMBL; M18381; LCVLA.
DR EMBL; M18382; LCVLB.
DR EMBL; M18383; LCVLC.
DR PIR; A30181; RXPIC.
KW RNA-DIRECTED RNA POLYMERASE.
FT CONFLICT 164 164 L -> Y (IN REF. 2).
FT CONFLICT 354 354 Q -> R (IN REF. 2).
FT CONFLICT 361 361 K -> E (IN REF. 2).
FT CONFLICT 382 382 H -> D (IN REF. 2).
FT CONFLICT 552 552 C -> S (IN REF. 2).
FT CONFLICT 1727 1727 R -> L (IN REF. 2).
SQ SEQUENCE 2210 AA; 254529 MW; 20447775 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXLX
|

NPKAFKSLKDLWDYMLNNTKGVLEFSI
1930 1940 X 1950

5. US-08-121-713B-30 (1-7)

BIME_EMENI NEGATIVE REGULATOR OF MITOSIS.

ID BIME_EMENI STANDARD; PRT; 2073 AA.
AC P24686;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE NEGATIVE REGULATOR OF MITOSIS.
GN BIME.
OC EMERICELLA NIDULANS (ASPERGILLUS NIDULANS).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; PLECTOMYCETES; EUROTIALES.
RN [1]
RP SEQUENCE FROM N.A.
RM 90375468
RA ENGLE D.B., OSMANI S.A., OSMANI A.H., ROSBOROUGH S., XIANG X.,
RA MORRIS N.R.;
RL J. BIOL. CHEM. 265:16132-16137(1990).
CC -1- FUNCTION: NEGATIVE REGULATOR OF MITOSIS IN E. NIDULANS. THIS
CC PROTEIN IS PART OF A REGULATORY PATHWAY THAT INCLUDES THE NIMA
CC PROTEIN KINASE. IT IS REQUIRED TO PREVENT PREMATURE ENTRY INTO
CC MITOSIS. MUTATIONS TO THIS PROTEIN BOTH CAUSE CELLS TO ENTER
CC MITOSIS AND PREVENT THEM FROM LEAVING MITOSIS.
DR EMBL; M59705; ANBIME.
DR PIR; A37879; BWASEE.
KW TRANSMEMBRANE; MITOSIS.
FT DOMAIN 342 353
FT TRANSMEM 1623 1643
FT TRANSMEM 1685 1703

FT TRANSEM 1746 1764 POTENTIAL.
SQ SEQUENCE 2073 AA; 229178 MW; 19117395 CN;
Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX
SSSGSYVDTSWESMLEQSSSAGWA
850 X 860 870

6. US-08-121-713B-30 (1-7)
CCG1_HUMAN TRANSCRIPTION INITIATION FACTOR TFIIID 250 KD SUBUN

ID CCG1_HUMAN STANDARD; PRT; 1872 AA.
AC P21675;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION INITIATION FACTOR TFIIID 250 KD SUBUNIT (TBP-ASSOCIATED
DE FACTOR 250 KD) (P250) (TAFII-250) (CELL CYCLE GENE 1 PROTEIN).
GN CCG1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LARYNGEAL CARCINOMA;
RM 91246200
RA SEKIGUCHI T., NOHIRO Y., NAKAMURA Y., HISAMOTO N., NISHIMOTO T.;
RL MOL. CELL. BIOL. 11:3317-3325(1991).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RM 89003056
RA SEKIGUCHI T., MIYATA T., NISHIMOTO T.;
RL EMBO J. 7:1683-1687(1988).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RM 93196704
RA RUPPERT S., WANG E.H., TJIAN R.;
RL NATURE 362:175-179(1993).
RN [4]
RP CHARACTERIZATION.
RM 93196705
RA HISATAKE K., HASEGAWA S., TAKADA R., NAKATANI Y., HORIKOSHI M.,
RA ROEDER R.G.;
RL NATURE 362:179-181(1993).
CC -!- FUNCTION: MAY PLAY AN ESSENTIAL ROLE IN TFIIID ASSEMBLY BY
CC INTERACTING WITH BOTH TBP AND OTHER TAF, AS WELL AS SERVING TO
CC LINK THE CONTROL OF TRANSCRIPTION TO THE CELL CYCLE. ESSENTIAL
CC FOR PROGRESSION OF THE G1 PHASE OF THE CELL CYCLE. POSSESSES
CC DNA-BINDING ACTIVITY.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SUBUNIT: TF2D IS COMPOSED OF TBP AND A VARIETY OF TBP-ASSOCIATED
CC FACTORS.
CC -!- PTM: PHOSPHORYLATED BY CASEIN KINASE II IN VITRO.
CC -!- SIMILARITY: CONTAINS TWO COPIES OF THE BROMODOMAIN.

-!- SIMILARITY: CONTAINS A HMG BOX.
CC DR EMBL; D90359; HSCCG1B.
DR EMBL; X07024; HSCCG1.
DR PIR; S03005; S03005.
DR PIR; A40262; A40262.
DR MIM; 313650; 11TH EDITION.
DR PROSITE; PS00633; BROMODOMAIN.
KW BROMODOMAIN; NUCLEAR PROTEIN; DNA-BINDING; CELL CYCLE; REPEAT;
KW TRANSCRIPTION REGULATION; PHOSPHORYLATION.
FT DOMAIN 157 165
FT DNA BIND 1195 1273 HMG BOX (POTENTIAL).
FT DOMAIN 1351 1358 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1402 1462 BROMODOMAIN.
FT DOMAIN 1525 1585 BROMODOMAIN.
FT DOMAIN 1627 1872 ASP/GLU-RICH (ACIDIC TAIL).
SQ SEQUENCE 1872 AA; 212676 MW; 16805423 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXXLX
VAEWRYGPARLWYDMLGVPEGSGFDY
340 350 X 360

7. US-08-121-713B-30 (1-7)
CALF_HUMAN COLLAGEN ALPHA 1(XVI) CHAIN PRECURSOR.

ID CALF_HUMAN STANDARD; PRT; 1603 AA.
AC Q07092;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE COLLAGEN ALPHA 1(XVI) CHAIN PRECURSOR.
GN COL16A1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RM 92335339
RA PAN T.C., ZHANG R.Z., MATTEI M.G., TIMPL R., CHU M.-L.;
RL PROC. NATL. ACAD. SCI. U.S.A. 89:6565-6569(1992).
RN [2]
RP SEQUENCE OF 421-1603 FROM N.A.
RC TISSUE=PLACENTA;
RA YANAGUCHI N., KIMURA S., MCBRIDE O.W., HORI H., YAMADA Y.,
RA KANAMORI T., YAMAKOSHI H., NAGAI Y.;
RL J. BIOCHEM. 112:856-863(1992).
CC -!- FUNCTION: THE NUMEROUS INTERRUPTIONS IN THE TRIPLE HELIX MAY MAKE
CC THIS MOLECULE EITHER ELASTIC OR FLEXIBLE.
CC -!- DOMAIN: THIS SEQUENCE DEFINES EIGHTEEN DIFFERENT DOMAINS, NINE
CC TRIPLE-HELICAL DOMAINS (COL9 TO COL1) AND TEN NONTRIPLE-HELICAL
CC DOMAINS (NC10 TO NC1).
CC -!- DEVELOPMENTAL STAGE: TRANSIENTLY ELEVATED EXPRESSION DURING
CC GESTATION, AND DECREASE AT TERM.
CC -!- TISSUE SPECIFICITY: IN THE PLACENTA, WHERE IT IS FOUND IN THE

CC AMNION, A MEMBRANOUS TISSUE LINING THE AMNIOTIC CAVITY. WITHIN THE
 CC AMNION, IT IS FOUND IN AN ACCELLULAR, RELATIVELY DENSE LAYER OF A
 CC COMPLEX NETWORK OF RETICULAR FIBRES. ALSO LOCATED TO A FIBROBLAST
 CC LAYER BENEATH THIS DENSE LAYER. EXISTS IN TISSUES IN ASSOCIATION
 CC WITH OTHER TYPES OF COLLAGEN.
 CC -|- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -|- SIMILARITY: BELONGS TO THE FAMILY OF FIBRIL-ASSOCIATED COLLAGENS
 CC WITH INTERRUPTED HELICES (FACIT).

DR EMBL; M92642; HSCOL16A.
 DR EMBL; S57132; S57132.
 DR PIR; S23810; S23810.
 DR MIM; 120326; 11TH EDITION.
 KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; COLLAGEN; HYDROXYLATION;
 KW REPEAT; SIGNAL.
 FT SIGNAL 1 21
 FT CHAIN 22 1603
 FT DOMAIN 22 374
 FT DOMAIN 375 505
 FT DOMAIN 506 520
 FT DOMAIN 521 554
 FT DOMAIN 555 571
 FT DOMAIN 572 630
 FT DOMAIN 631 651
 FT DOMAIN 652 722
 FT DOMAIN 723 737
 FT DOMAIN 738 875
 FT DOMAIN 876 886
 FT DOMAIN 887 938
 FT DOMAIN 939 972
 FT DOMAIN 973 987
 FT DOMAIN 988 1010
 FT DOMAIN 1011 1432
 FT DOMAIN 1433 1471
 FT DOMAIN 1472 1577
 FT DOMAIN 1578 1603
 FT CONFLICT 537 537
 FT CONFLICT 1165 1165
 FT CONFLICT 1603 AA; 157692 MW; 12917962 CN;
 SQ SEQUENCE 1603 AA; 157692 MW; 12917962 CN;
 Initial Score = 7 Optimized Score = 7 Significance = 3.99
 Residue Identity = 28% Matches = 7 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X X
 X XXXXIX
 MWVSWAPCLWLLGLWATFGHANTG
 10 X 20

8. US-08-121-713B-30 (1-7)

PHLX_RABIT PHOSPHOLIPASE ADRAB-B PRECURSOR (EC 3.1.-.-).
 ID PHLX_RABIT STANDARD; PRT; 1458 AA.
 AC Q05017;
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE PHOSPHOLIPASE ADRAB-B PRECURSOR (EC 3.1.-.-).
 OS ORYCTOLAGUS CONICULOSUS (RABBIT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; LAGOMORPHA.
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=INTESTINE;
 RM 93286138

RA BOLL W., SCHMID-CHANDA T., SEMENZA G., MANTEI N.;
 RL J. BIOL. CHEM. 268:12901-12911 (1993).
 CC -|- FUNCTION: HAS ESTERASE AND PHOSPHOLIPASE A/LYSOPHOSPHOLIPASE
 CC ACTIVITY. CAN CONVERT PHOSPHATIDYLCHOLINE TO FATTY ACIDS AND
 CC GLYCEROPHOSPHOCHOLINE. COULD BE INVOLVED IN UPTAKE OF DIETARY
 CC LIPIDS, POSSIBLY INCLUDING LONG CHAIN RETINYL ESTERS.
 CC -|- DEVELOPMENTAL STAGE: EXPRESSED IN THE INTESTINE OF ADULT BUT NOT
 CC BABY RABBITS.
 CC -|- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. BRUSH BORDER.
 CC -|- TISSUE SPECIFICITY: INTESTINE.
 DR EMBL; 212841; OCPHPLIP.
 DR PIR; A45665; A45665.
 KW HYDROLASE; REPEAT; SIGNAL.
 FT SIGNAL 1 19
 FT CHAIN 20 1458
 FT DOMAIN 20 1415
 FT TRANSMEM 1416 1439
 FT DOMAIN 1440 1458
 FT DOMAIN 39 1403
 FT REPEAT 39 347
 FT REPEAT 362 707
 FT REPEAT 708 1054
 FT REPEAT 1064 1403
 SQ SEQUENCE 1458 AA; 161343 MW; 10794453 CN;
 Initial Score = 7 Optimized Score = 7 Significance = 3.99
 Residue Identity = 28% Matches = 2 Mismatches = 5
 Gaps = 0 Conservative Substitutions = 0

X X X
 X XXXXIX
 SKAHAHAASALNNMPEVPGOKTTTHND
 660 670 X 680

9. US-08-121-713B-30 (1-7)
 EBAL_PLAFC ERYTHROCYTE-BINDING ANTIGEN EBA-175.

ID EBAL_PLAFC STANDARD; PRT; 1426 AA.
 AC P19214;
 DT 01-NOV-1990 (REL. 16, CREATED)
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
 DT 01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)
 DE ERYTHROCYTE-BINDING ANTIGEN EBA-175.

OS PLASMODIUM FALCIPARUM (ISOLATE CAMP / MALAYSIA).
OC EUKARYOTA; PROTOZOA; APICOMPLEXA; SPOROZOA; COCCIDIA; EUCCOCCIDIIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LATE SCHIZONT;
RM 90377299
RA SIM B.K.L.;
RL MOL. BIOCHEM. PARASITOL. 41:293-296 (1990).
DR EMBL; X52524; PFEBA175.
DR PIR; S11561; S11561.
KW ANTIGEN.
FT DOMAIN 159 1101
FT VARIANT 1028 1028
SQ SEQUENCE 1426 AA; 166155 MW; 10311930 CN;

ESSENTIAL FOR BINDING TO
ERYTHROCYTES.
E -> V (IN STRAINS PCR3 AND ITG).
FT VARIANT 1028 1028
SQ SEQUENCE 1426 AA; 166155 MW; 10311930 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXLX
KERNREKLEWAMLSHKNINCK
290 X 300 310

10. US-08-121-713B-30 (1-7)
KRE5_YEAST KILLER TOXIN-RESISTANCE PROTEIN 5 PRECURSOR.

ID KRE5_YEAST STANDARD; PRT; 1365 AA.
AC P22023;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)
DE KILLER TOXIN-RESISTANCE PROTEIN 5 PRECURSOR.
GN KRE5.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RM 90258892
RA MEADEN P., HILL K., WAGNER J., SLIPETZ D., SOMMER S.S., BUSSEY H.;
RL MOL. CELL. BIOL. 10:3013-3019 (1990).
CC -1- FUNCTION: REQUIRED FOR (1->6)-BETA-D-GLUCAN SYNTHESIS AND NORMAL
CELL GROWTH.
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.

DR EMBL; M33536; SKRFS.
DR PIR; S12202; BVBYK5.
DR LISTA; SC00535; KRE5.
DR PROSITE; PS00014; ER TARGET.
KW CELL WALL; ENDOPLASMIC RETICULUM; SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 17
FT CHAIN 18 1365
FT CARBOHYD 115 115
FT CARBOHYD 228 228
FT CARBOHYD 293 293
FT CARBOHYD 457 457
FT CARBOHYD 519 519
FT CARBOHYD 523 523

POTENTIAL.
KILLER TOXIN-RESISTANCE PROTEIN 5.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.

FT CARBOHYD 643 643 POTENTIAL.
FT CARBOHYD 789 789 POTENTIAL.
FT CARBOHYD 870 870 POTENTIAL.
FT CARBOHYD 1091 1091 POTENTIAL.
FT CARBOHYD 1150 1150 POTENTIAL.
FT CARBOHYD 1195 1195 POTENTIAL.
FT SITE 1362 1365 PREVENT SECRETION FROM ER.
SQ SEQUENCE 1365 AA; 156484 MW; 9856080 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXLX
MSGDGYWKEGYWEKMLRNNLEFYSTE
1230 1240 1250

11. US-08-121-713B-30 (1-7)
YB95_YEAST HYPOTHETICAL 132.7 KD HELICASE IN ALG7-HIS7 INTERG

ID YB95_YEAST STANDARD; PRT; 1143 AA.

AC P38144;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 132.7 KD HELICASE IN ALG7-HIS7 INTERGENIC REGION.
GN YB245C OR YBR1633.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA ALJINOVIC G., POHL F.M., POHL T.M.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO HELICASES OF THE SNF2/RAD54 FAMILY.
DR EMBL; Z36114; SCYBR245C.
DR PIR; S46122; S46122.
KW HYPOTHETICAL PROTEIN; NUCLEAR PROTEIN; DNA-BINDING; HELICASE;
KW ATP-BINDING.

FT NP BIND 235 242 ATP (BY SIMILARITY).
FT SITE 338 341 DEGH BOX.
SQ SEQUENCE 1143 AA; 132660 MW; 6451971 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
XWXXLX
PARGGCRNTLLPNNMLAIFANFHFFKY
10 X 20 30

12. US-08-121-713B-30 (1-7)
SP23_YEAST SPT23 PROTEIN.

ID SP23 YEAST STANDARD; PRT; 1082 AA.
AC P35210;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE SPT23 PROTEIN.
GN SPT23 OR YK1020C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN RIEGER M.;
RP SEQUENCE FROM N.A.
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RM BURKETT T.J., GARFINKEL D.J.;
RA YEAST 10:81-92(1994).
CC -1- FUNCTION: DOSAGE-DEPENDENT SUPPRESSOR OF TY-INDUCED PROMOTER
CC MUTATIONS. MAY EXERT IT SUPPRESSION EFFECT THROUGH PROTEIN-PROTEIN
CC INTERACTIONS SINCE DOES NOT PRESENT ANY OF THE MOTIFS GENERALLY
CC FOUND IN TRANSCRIPTIONAL ACTIVATORS OR DNA BINDING PROTEINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 738
CC ONWARD AND IS SHORTER (752 AA) DUE TO A FRAMESHIFT.
CC -1- SIMILARITY: TO YEAST YI033W.
DR EMBL; 228020; SCYK1020C.
DR EMBL; L24760; SCSEPT23A.
DR PIR; S37837; S37837.
DR LISTA; SC01182; SPT23.
FT CONFLICT 715 715 H -> P (IN REF. 2).
SQ SEQUENCE 1082 AA; 121337 MW; 5973945 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

TKTSEIDPGSLWNPMLTRNDELPKYE
920 930 X 940

13. US-08-121-713B-30 (1-7)
PPS3_BACSU PEPTIDE SYNTHETASE 3 (FRAGMENT).
ID PPS3 BACSU STANDARD; PRT; 859 AA.
AC P39847;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE PEPTIDE SYNTHETASE 3 (FRAGMENT).
GN BACILLUS SUBTILIS.
OS PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA TOGNONI A., GRANDI G.;

RL SUBMITTED (JUN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
DR EMBL; Z34883; BSPEPSYN.
DR SUBTILIST; BG10972; PPSC.
KW MULTIFUNCTIONAL ENZYME; LIGASE; REPEAT.
FT NON TER 859 859
SQ SEQUENCE 859 AA; 96679 MW; 3839344 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X X
XWXXXLX
GATINTVFQALWGMICOKVNSSDAVF
250 260 X 270

14. US-08-121-713B-30 (1-7)
YLH5_CAEEL PUTATIVE ABC TRANSPORTER C48B4.5 IN CHROMOSOME III
ID YLH5 CAEEL STANDARD; PRT; 856 AA.
AC P34359;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE PUTATIVE ABC TRANSPORTER C48B4.5 IN CHROMOSOME III.
GN C48B4.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RM 94150718
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FRASER A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RL NATURE 368:32-38(1994).
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS).
DR EMBL; 229117; CEC48B4.
DR PIR; S40725; S40725.
DR WORMPEP; C48B4.5; CE00488.
DR PROSITE; PS00211; ABC TRANSPORTER.
KW HYPOTHETICAL PROTEIN; ATP-BINDING; TRANSPORT.
FT NP BIND 563 570 ATP (POTENTIAL).
SQ SEQUENCE 856 AA; 95999 MW; 3881360 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5

Gaps = 0 Conservative Substitutions = 0

X X
X XXXXLX
|

AGIDPKARREVWELLWCREHSNSALM
690 X 700

15. US-08-121-713B-30 (1-7)
SYFB_BACSU PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.1.

ID SYFB_BACSU STANDARD; .PRT; 804 AA.
AC P17922;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLALANINE--
DE TRNA LIGASE BETA CHAIN).
GN PHET.
OS BACILLUS SUBTILIS.
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RM 91175935
RA BRAKHAGE A., WOZNY M., PUTZER H.;
RL BIOCHIMIE 72:725-734(1990).
RN [2]
RP ERRATUM.
RM 91234765
RA BRAKHAGE A., WOZNY M., PUTZER H.;
RL BIOCHIMIE 73:127-127(1991).
CC -!- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) = AMP +
CC -!- PYROPHOSPHATE + L-PHENYLALANYL-TRNA(PHE).
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
DR EMBL; X53057; BSPHEST.
DR PIR; S11731; YFBSB.
DR SUBTILIST; BG10875; PHET.
KW AMINOACYL-TRNA SYNTHASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
SQ SEQUENCE 804 AA; 87917 MW; 3398444 CN;

Initial Score = 7 Optimized Score = 7 Significance = 3.99
Residue Identity = 28% Matches = 2 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
X XXXXLX
|

TERVAGAVTGLWRKQLWQGEKPVDF
600 610 X 620

maryh@stic

stdin

NeWSprinter20

Fri May 19 11:05:07 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5

7. R55706 Galactosyltransferase.
 8. R28838 HeLa cell galactosyltransferase
 9. R05932 Human beta-1,4-galactosyltransferase
 10. R05932 Mouse beta-1,4-galactosyltransferase
 11. R55130 Rice alpha-amylase coding.
 12. R32987 Rice alpha-amylase.
 13. R10694 Cephalosporin antibiotic biosynthesis
 14. R42078 Impatiens Necrotic Spot Virus
 15. R34135 C. roseus TDC with mutated N-terminus
 16. R34133 Catharanthus roseus tryptophan
 17. R06829 Tryptamine.
 18. R44139 Human FACC.
 19. R06520 Microspore-specific clone L19
 20. R43382 Plant NADPH cytochrome P450 r
 21. R5709 Glycosyltransferase hybrid.
 22. R5708 Glycosyltransferase hybrid.
 23. R04197 Env gene of simian immunodeficiency
 24. R90954 Yeast topoisomerase I cDNA
 25. R47519 HIV-4 ribonucleotide reductase
 26. R25589 RING4 antigenic peptide transmembrane
 27. R45749 Alpha-DNA polymerase.
 28. R43996 Thermophilic DNA polymerase.
 29. R61082 Complete translation of plasmid
 30. R61030 Entire coded sequence from plasmid
 31. R61036 Translation of plasmid pAU157
 32. R44401 Canine coronavirus 1-71 spike
 33. R5764 Sequence encoded by the cDNA
 34. R41668 Human endothelial cell nitric
 35. R38698 S-PRV-055 TGE virus gp195 gene
 36. R27818 CCV-6 spike protein.
 37. R27819 CCVinsavc spike protein.
 38. R42471 Canine coronavirus 1-71 spike
 39. R44400 Canine coronavirus 1-71 spike
 40. R31038 E2 protein of Canine Corona V
 41. R27820 CCV-C54 spike protein.
 42. R42478 FECV/FIPV chimeric spike prot
 43. R42477 FECV/FIPV chimeric spike prot
 44. R42475 FECV/FIPV chimeric spike prot
 45. R42474 FECV/FIPV chimeric spike prot

1. US-08-121-713B-33 (1-9)
 R04198 F gene of simian immunodeficiency virus.

ID R04198 standard; protein; 95 AA.
 AC R04198;
 DT 12-JUN-1990 (first entry)
 DE F gene of simian immunodeficiency virus.
 KW SIV; simian immunodeficiency virus; AIDS; HIV; vaccine; ss.
 OS SIV.
 PN J01289485-A.
 PD 21-NOV-1989.
 PF 16-JUN-1988; 119023.
 PA (TOFU) Toa Nenryo Kogyo KK.
 DR WPI; 90-005176/01.
 DR N-PSDB; Q02829.
 PT DNA complementary to RNA of simian immunodeficiency virus -
 PT used in vaccines and for diagnosis of AIDS.
 PS Disclosure; Fig 4; 6pp; Japanese.

CC F gene derived from RNA of SIV from which vector plasmid pSAH121
 CC may be constructed and incorporated into an E.coli vector. Useful in
 CC developing a vaccine against and in diagnosis of AIDS.
 SQ Sequence 95 AA;
 SQ 9 A; 8 R; 6 N; 1 D; 0 B; 4 C; 9 Q; 4 E; 0 Z; 4 G; 1 H;
 SQ 5 L; 16 L; 4 K; 1 M; 1 F; 1 P; 5 S; 6 T; 3 W; 4 Y; 3 V;
 Initial Score = 7 Optimized Score = 7 Significance = 3.45
 Residue Identity = 33% Matches = 3 Mismatches = 6
 Gaps = 0 Conservative Substitutions = 0

X X
 WXXIXKXXL
 SKSRGCMQLTAMITRLNTWLYNSCLTLLIQ
 10 X 20 30

2. US-08-121-713B-33 (1-9)
 R51074 Sphi restriction endonuclease.

ID R51074 standard; Protein; 235 AA.
 AC R51074;
 DT 24-MAY-1994 (first entry)
 DE Sphi restriction endonuclease.
 KW Restriction endonuclease; methylase; Sphi; fermentation;
 QS Streptomyces phaeochromogenes.
 PN DS5262318-A.
 PD 16-NOV-1993.
 PF 20-AUG-1992; 932454.
 PR 20-AUG-1992; US-932454.
 PA (NEW) NEW ENGLAND BIOLABS INC.
 PI Guthrie EP;
 DR WPI; 93-377440/47.
 DR N-PSDB; Q51114.
 PT Sphi restriction endo-nuclease prepn. - comprises cloning Sphi
 PS Example 1; Columns 25-26; 20pp; English.
 CC A DNA sequence comprising the methylase and Sphi restriction
 CC endonuclease genes (Q51114) is used in the construction of a
 CC recombinant vector. The vector is then used to transform E. coli
 CC RRI which can then express the restriction endonuclease gene to
 CC produce the enzyme. The restriction endonuclease gene is cloned
 CC alongside the relevant methylase gene to prevent degradation of
 CC the host cell genome when the restriction endonuclease is produced.
 SQ Sequence 235 AA;
 SQ 29 A; 17 R; 9 N; 17 D; 0 B; 8 C; 5 Q; 16 E; 0 Z; 17 G; 4 H;
 SQ 8 I; 18 L; 17 K; 5 M; 8 F; 11 P; 13 S; 9 T; 4 W; 6 Y; 14 V;

Initial Score = 7 Optimized Score = 7 Significance = 3.45
 Residue Identity = 33% Matches = 3 Mismatches = 6
 Gaps = 0 Conservative Substitutions = 0

X X
 WXXIXKXXL
 PIVLSADQIAWLRQIKMSKRAALVRDYIL
 10 X 20 X 30

3. US-08-121-713B-33 (1-9)
R05934 Bovine beta-1,4-galactosyltransferase.

ID R05934 standard; protein; 331 AA.
AC R05934;
DT 22-NOV-1990 (first entry)
DE Bovine beta-1,4-galactosyltransferase.
KW Beta-1,4-galactosyl transferase; sugars; glycoproteins;
KW glycolipids; dyserythropoietic anaemia type II; ds.
OS Bos taurus.
PN W09007000-A.
PD 28-JUN-1990.
PF 16-NOV-1989; U05128.
PR 13-DEC-1988; US-283732.
PA (JOLL-) LA JOLLA CANCER RES.
PI FUKUDA MN, APPERT HA;
DR WPI; 90-224528/29.
PT New nucleic acid sequences encoding beta-1,4-galactosyl
PT transferase - in bound or soluble forms, and derived
PT polypeptide(s) and antibodies, useful in synthesis and diagnosis.
PS Disclosure; p; English.
CC Derived non-membrane bound peptide product can be collected
CC as a cell secretion from a transformed host. Product is useful
CC in production of sugars, glycoproteins and glycolipids. Abs
CC raised to the product may be used in diagnosis of abnormal
CC conditions such as congenital dyserythropoietic anaemia type II.
SQ Sequence 331 AA;
SQ 14 A; 20 R; 19 N; 17 D; 0 B; 5 C; 16 Q; 0 Z; 25 G; 8 H;
SQ 20 I; 32 L; 15 K; 9 M; 13 F; 30 P; 24 S; 13 T; 3 W; 18 Y; 18 V;
Initial Score = 7 Optimized Score = 7 Significance = 3.45
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXXLKXXL
KVATIPFRNRQEHKVMYLYLHPILORQ
110 120 X 130

4. US-08-121-713B-33 (1-9)
R31687 MVP-5180/91 fragment (gp120/gp41).

ID R31687 standard; protein; 351 AA.
AC R31687;
DT 07-NOV-1994 (first entry)
DE MVP-5180/91 fragment (gp120/gp41).
KW Human immunodeficiency virus; HIV; antigen; detection; diagnosis;
KW retrovirus; vaccine; lymphocyte; reverse transcriptase;
KW amplification; primer; polymerase chain reaction; PCR.
OS Synthetic.
FH Key Location/Qualifiers
FT Region 1..16
FT /label= gp120
FT Region 17..351
FT /label= gp41
PN EP-591914-A.
PD 13-APR-1994.

PF 05-OCT-1993; 116058.
PR 06-OCT-1992; DE-233646.
PR 22-OCT-1992; DE-235718.
PR 30-DEC-1992; DE-244541.
PR 01-JUN-1993; DE-318186.
PA (BEHW) BEHRINGER AG.
PI Brunn V. A, Eberle J, Gurtler LG, Hauser H-P, Knapp S;
DR WPI; 94-120077/15.
DR N-PSDB; Q38966.
PT New HIV-type immune deficiency virus ECACC V 92092318 - and
PT deriv. cDNA or antigens, useful for diagnosing retroviral
PT infections and vaccines.
PS Example 7; Page 13-14; 73pp; German.
CC MVP-5180/91 DNA is obtained by PCR using the primers given
CC in Q58962-65. The primer given in Q58963 is derived from HIV-1
CC BH10 (bases 8129-8109).
CC The obtained fragment is given in Q58966.
SQ Sequence 351 AA;
SQ 24 A; 20 R; 17 N; 14 D; 0 B; 4 C; 29 Q; 13 E; 0 Z; 28 G; 7 H;
SQ 36 I; 49 L; 13 K; 5 M; 3 F; 9 P; 17 S; 20 T; 17 W; 9 Y; 17 V;
Initial Score = 7 Optimized Score = 7 Significance = 3.45
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXXLKXXL
CRLCGAVMOYWOELKNSATNLDITAVS
310 X 320 X 330

5. US-08-121-713B-33 (1-9)
R44211 Caffeine demethylase.

ID R44211 standard; protein; 351 AA.
AC R44211;
DT 26-MAY-1994 (first entry)
DE Caffeine demethylase.
KW Caffeine demethylase; CDM; Pseudomonas; 3-methyl-7-alkyl-xanthine;
KW 1,3-dimethyl-7-alkyl-xanthine; demethylation; ds.
OS Pseudomonas sp.
PN DE4316882-A.
PD 25-NOV-1993.
PF 19-MAY-1993; 316882.
PR 20-MAY-1992; JP-154380.
PR 27-OCT-1992; JP-312954.
PA (AMANO) AMANO PHARM KK.
PI Imai Y, Koide Y, Nakane S;
DR WPI; 93-378610/48.
DR N-PSDB; Q51766.
PT DNA contg. caffeine de-methylase gene - used in prodn. of
PT 3-methyl-7-alkyl-xanthine cpds. by microbial demethylation of
PT 1,3-di:methyl derivs.
PS Claim 4; Page 16-17; 23pp; German.
CC 3-methyl-7-alkyl-xanthine cpds. can be produced by culturing
CC a transformant contg. the CDM gene in a nutrient contg.
CC 1,3-dimethyl-7-alkyl-xanthine.
CC 3-methyl-7-alkyl-xanthine cpds. are useful as pharmaceutical
CC intermediates.

SQ Sequence 351 AA;
 SQ 31 A; 19 R; 17 N; 26 D; 0 B; 9 C; 14 Q; 22 E; 0 Z; 11 G; 11 H;
 SQ 15 I; 26 L; 14 K; 5 M; 19 F; 24 P; 27 S; 14 T; 8 W; 14 Y; 25 V;
 Initial Score = 7 Optimized Score = 7 Significance = 3.45
 Residue Identity = 33% Matches = 3 Mismatches = 6
 Gaps = 0 Conservative Substitutions = 0

X X
 WXXLXXXXL
 I I
 ADVKSIOYRWLRELKEAHQDGAQAFRSA
 310 320 330

6. US-08-121-713B-33 (1-9)
 P70668 D-alanine racemase.

ID P70668 standard; protein; 389 AA.
 AC P70668;
 DT 11-MAR-1991 (first entry)
 DE D-alanine racemase.
 KW L-glutamate racemase; D-alanyl-D-alanine ligase.
 OS Bacillus subtilis.
 PN GB2177097-A.
 PD 14-JAN-1987.
 PF 17-JUN-1986; 014702.
 PR 18-JUN-1985; US-746437.
 PA (GENE-) GENENCOR INC.
 PI Ferrari E;
 DR WPI; 87-009482/02.
 DR N-PSDB; N70668.

PT Stable maintenance of heterologous DNA in host cell - using host
 PT cells deficient in ability to synthesise cell wall transformed to
 PT regain ability
 PS Disclosure; Fig 2; 13pp; English.
 CC Protein may be produced in a transformed cell-line deficient in
 CC D-alanine production. The cell will then express the required sequence
 CC and a second sequence operably linked to the D-alanine racemase.
 CC Other sequences which may be used in the same way include those
 CC encoding L-glutamate racemase or D-alanyl-D-alanine ligase.

SQ Sequence 389 AA;
 SQ 37 A; 21 R; 14 N; 18 D; 0 B; 5 C; 8 Q; 29 E; 0 Z; 29 G; 11 H;
 SQ 23 I; 38 L; 30 K; 15 M; 14 F; 15 P; 19 S; 20 T; 4 W; 13 Y; 26 V;

Initial Score = 7 Optimized Score = 7 Significance = 3.45
 Residue Identity = 33% Matches = 3 Mismatches = 6
 Gaps = 0 Conservative Substitutions = 0

X X
 WXXLXXXXL
 I I
 GTVPVGYADGWLRLKLGTDILVKGRIKI
 280 290 300

7. US-08-121-713B-33 (1-9)
 R55706 Galactosyltransferase.
 ID R55706 standard; Protein; 398 AA.

AC R55706;
 DT 18-NOV-1994 (first entry)
 DE Galactosyltransferase.
 KW Galactosyltransferase; sialyltransferase; hybrid protein;
 KW glycosyltransferase; glycoprotein; glycolipid; oligosaccharide;
 KW Hela; Escherichia coli; pAD113; glycosylation;
 KW Saccharomyces cerevisiae.
 OS Homo sapiens.
 PN WO9412646-A.
 PD 09-JUN-1994.
 PF 15-NOV-1993; E03194.
 PR 27-NOV-1992; EP-810924.
 PA (CIBA) CIBA GEIGY AG.
 PI Berger EG, Iwanow SX, Watzele M;
 DR WPI; 94-200274/24.
 DR P-PSDB; Q66889.

PT Proteins with glycosyl transferase activity - useful for
 PT synthesis or modification of glyco-proteins, glyco-lipid(s) and
 PT oligosaccharide(s)
 PS Disclosure; Page 32-34; 67pp; English.
 CC cDNA encoding galactosyltransferase (GT) was isolated from Hela
 CC cells, cloned by PCR using the primers given in Q66893-97, and
 CC expressed in Escherichia coli DH5-alpha using plasmid pAD113.
 CC Hybrid proteins (R55708, R55709) comprising membrane-bound or
 CC soluble derivatives of GT linked to soluble sialyltransferase
 CC were prepared in Saccharomyces cerevisiae.
 SQ Sequence 398 AA;
 SQ 26 A; 27 R; 19 N; 21 D; 0 B; 7 C; 20 Q; 11 E; 0 Z; 33 G; 9 H;
 SQ 17 I; 42 L; 11 K; 10 M; 15 F; 35 P; 30 S; 15 T; 3 W; 17 Y; 30 V;

Initial Score = 7 Optimized Score = 7 Significance = 3.45
 Residue Identity = 33% Matches = 3 Mismatches = 6
 Gaps = 0 Conservative Substitutions = 0

X X
 WXXLXXXXL
 I I
 KVAIIIPFNROEHLKWLWLYLHPVLQKQ
 180 190 X 200

8. US-08-121-713B-33 (1-9)
 R28838 Hela cell galactosyltransferase enzyme.

ID R28838 standard; Protein; 398 AA.
 AC R28838;
 DT 03-APR-1993 (first entry)
 DE Hela cell galactosyltransferase enzyme.
 KW glycosyltransferase; galactosyltransferase; sialyltransferase;
 KW fucosyltransferase; membrane bound; ss.
 OS Homo sapiens.
 PN GB2256197-A.
 PD 02-DEC-1992.

PF 14-APR-1991; 008211.
 PR 31-MAY-1991; EP-810414.
 PR 04-MAR-1992; EP-810167.
 PR 14-APR-1992; GB-008211.
 PA (CIBA) CIBA GEIGY AG.
 PI Berger EG, Meyhack B, Watzele G, Watzele M, Berger E;
 DR WPI; 92-401159/49.

DR N-PSDB; Q31434.
PT Glycosyltransferase prodn. process - includes transforming yeast
PT cells with expression cassettes contg. mammalian coding sequences
PT controlled by yeast promoters
PS Claim 7; Page 43; 65pp; English.
CC This sequence represents a galactosyltransferase enzyme from HeLa cells.
CC (EC 2.4.1.22). It was decoded from the appropriate cDNA. The method
CC of the invention is used to produce membrane-bound mammalian
CC glycosyltransferase and variants, using transformed yeasts. It is
CC less time consuming than natural source isolation and chemical methods.
SQ Sequence 398 AA;
SQ 26 A; 27 R; 19 N; 21 D; 0 B; 7 C; 20 Q; 11 E; 0 Z; 33 G; 9 H;
SQ 17 I; 42 L; 11 K; 10 M; 15 F; 34 P; 31 S; 15 T; 3 W; 17 Y; 30 V;

Initial Score = 7 Optimized Score = 7 Significance = 3.45
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X X
WXXXLKXXL
||
KVAIIIFRNROEHLKYWLYLHPVLQRQ
180 190 X 200

9. US-08-121-713B-33 (1-9)
R05932 Human beta-1,4-galactosyltransferase.

ID R05932 standard; protein; 400 AA.
AC R05932;
DT 22-NOV-1990 (first entry)
DE Human beta-1,4-galactosyltransferase.
KW Beta-1,4-galactosyl transferase; sugars; glycoproteins;
KW glycolipids; dyserythropoietic anaemia type II; ds.

OS Homo sapiens.
FH Key Location/Qualifiers
FT Active-site 112..114
FT /label=Potential glycosylation site.
PN W09007000-A.

PD 28-JUN-1990. U05128.
PF 16-NOV-1989; US-283732.
PR 13-DEC-1988; US-283732.

PA (JOLL-) LA JOLLA CANCER RES.
PI FUKUDA MN, APPERT HA;
DR WPI; 90-224528/29.
DR N-PSDB; Q05265.

PT New nucleic acid sequences encoding beta-1,4-galactosyl
PT transferase - in bound or soluble forms, and derived
PT polypeptide(s) and antibodies, useful in synthesis and diagnosis.
PS Claim 3; Fig 2; 31pp; English.

CC Derived non-membrane bound peptide product can be collected
CC as a cell secretion from a transformed host. Product is useful
CC in production of sugars, glycoproteins and glycolipids. Abs
CC raised to the product may be used in diagnosis of abnormal
CC conditions such as congenital dyserythropoietic anaemia type II.

SQ Sequence 400 AA;
SQ 26 A; 27 R; 19 N; 21 D; 0 B; 7 C; 20 Q; 11 E; 0 Z; 34 G; 9 H;
SQ 18 I; 42 L; 11 K; 10 M; 15 F; 34 P; 30 S; 15 T; 3 W; 18 Y; 30 V;

Initial Score = 7 Optimized Score = 7 Significance = 3.45

Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXXLKXXL
||
KVAIIIFRNROEHLKYWLYLHPVLQRQ
180 X 190 X 200

10. US-08-121-713B-33 (1-9)
R05933 Mouse beta-1,4-galactosyltransferase.

ID R05933 standard; protein; 402 AA.
AC R05933;
DT 22-NOV-1990 (first entry)
DE Mouse beta-1,4-galactosyltransferase.
KW Beta-1,4-galactosyl transferase; sugars; glycoproteins;
KW glycolipids; dyserythropoietic anaemia type II; ds.

OS Mus sp.
PN W09007000-A.
PD 28-JUN-1990.
PF 16-NOV-1989; U05128.

PR 13-DEC-1988; US-283732.
PA (JOLL-) LA JOLLA CANCER RES.
PI FUKUDA MN, APPERT HA;
DR WPI; 90-224528/29.

PT New nucleic acid sequences encoding beta-1,4-galactosyl
PT transferase - in bound or soluble forms, and derived
PT polypeptide(s) and antibodies, useful in synthesis and diagnosis.
PS Disclosure; P; English.

CC Derived non-membrane bound peptide product can be collected
CC as a cell secretion from a transformed host. Product is useful
CC in production of sugars, glycoproteins and glycolipids. Abs
CC raised to the product may be used in diagnosis of abnormal
CC conditions such as congenital dyserythropoietic anaemia type II.

SQ Sequence 402 AA;
SQ 25 A; 29 R; 19 N; 22 D; 0 B; 7 C; 19 Q; 10 E; 0 Z; 35 G; 8 H;
SQ 20 I; 43 L; 18 K; 9 M; 17 F; 34 P; 27 S; 15 T; 3 W; 18 Y; 24 V;

Initial Score = 7 Optimized Score = 7 Significance = 3.45
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXXLKXXL
||
KVAIIIFRNROEHLKYWLYLHPVLQRQ
180 190 X 200

11. US-08-121-713B-33 (1-9)
R55130 Rice alpha-amylase coding.

ID R55130 standard; Protein; 433 AA.

AC R55130;
DT 12-JAN-1995 (first entry)
DE Rice alpha-amylase coding.
KW Virus; recombination; plant virus; alpha trichosanthin; phenotype;
KW alpha amylase; alpha haemoglobin; brome mosaic virus; gemini virus;

KW rice necrosis virus tobamovirus; gene expression; chinese cucumber.
OS Oryza sativa.
PN US5316931-A.
PD 31-MAY-1994. 160766.
PF 26-FEB-1988; US-160766.
PR 26-FEB-1988; US-160771.
PR 26-FEB-1988; US-219279.
PR 15-JUL-1988; US-310881.
PR 17-FEB-1989; US-347637.
PR 05-MAY-1989; US-347637.
PR 08-JUN-1989; US-363138.
PR 22-OCT-1990; US-600244.
PR 16-JAN-1991; US-641617.
PR 26-JUL-1991; US-737899.
PR 01-AUG-1991; US-739143.
PA (BIOS-) BIOSOURCE GENETICS CORP.
PI Dawson WO, Donson J, Garger SJ, Grantham GL, Grille IK;
PI Turpen AM, Turpen TH;
DR WPI; 94-176269/21.
DR N-PSDB; Q65574.
PT New recombinant plant viral nucleic acid - capable of systemic
PT infection and stable expression of non-native nucleic acid in
PT plant host
PS Example 4; Columns 53-56; 44pp; English.
CC The rice alpha-amylase gene may be inserted into a recombinant plant
CC virus which can then be used to infect plants for the production of
CC non-native products (in this case alpha-amylase). Other genes which
CC may be inserted into the virus are those which control a phenotypic
CC trait, such as male sterility, or sequences encoding anti-sense RNA
CC which can be useful to prevent the expression of undesired phenotypic
CC traits. The recombinant virus is derived from a plus sense, single
CC stranded virus selected from tobamovirus, brome mosaic virus, rice
CC necrosis virus or a gemini virus.
SQ Sequence 433 AA;
SQ 33 A; 17 R; 20 N; 36 D; 0 B; 3 C; 13 Q; 23 E; 0 Z; 47 G; 15 H;
SQ 29 I; 34 L; 20 K; 11 M; 18 F; 21 P; 20 S; 18 T; 15 W; 15 Y; 25 V;
Initial Score = 7 Optimized Score = 7 Significance = 3.45
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

NKRVQRELIGWLDWLKMDIGFDWRLDFA
190 X 200 210

12. US-08-121-713B-33 (1-9)
R32987 Rice alpha-amylase.

ID R32987 standard; Protein; 434 AA.
AC R32987;
DT 17-JUN-1993 (first entry)
DE Rice alpha-amylase.
KW Recombinant products; commercial production; fermentation;
KW biosynthesis; natural products; recombinant proteins;
KW product expression; protein expression; expressed proteins.
OS Oryza sativa.
PN WO9303161-A.

PD 18-FEB-1993.
PF 31-JUL-1992; U06359.
PR 01-AUG-1991; US-739143.
PA (DAMS/) DAWSON W O.
PA (DONS/) DONSON J.
PA (GARG/) GARGER S J.
PA (GRAN/) .
PA GRANTHAM G L.
PA (GRIL/) GRILLE L K.
PA (TURP/) TURPEN A M.
PA (TURP/) .
PA TURPEN T H.
PA Dawson WO, Donson J, Garger SJ, Grantham GL, Grille IK;
PI Turpenam, Turpen TH;
PI WPI; 93-076518/09.
DR N-PSDB; Q37680.
DR Recombinant plant viral nucleic acids - used to express a prod.,
PT e.g. antibody or IL-1 in a plant
PT Example 4; Page 96; 30pp; English.
PS This sequence is rice alpha amylase. The coding sequence was inserted
CC into a recombinant plant viral nucleic acid which was then used to
CC express a recombinant product (in this case rice alpha-amylase) in a
CC plant. The plant viral sequence may be from tobacco mosaic, cucumber
CC green mottle, cowpea mosaic, brome mosaic, broad bean mottle, rice
CC necrosis, geminiviruses, tomato golden mosaic, Cassava latent and
CC maize streak viruses.
SQ Sequence 434 AA;
SQ 34 A; 17 R; 20 N; 36 D; 0 B; 3 C; 13 Q; 23 E; 0 Z; 47 G; 16 H;
SQ 29 I; 34 L; 20 K; 11 M; 18 F; 20 P; 19 S; 18 T; 16 W; 15 Y; 25 V;
Initial Score = 7 Optimized Score = 7 Significance = 3.45
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

NKRVQRELIGWLDWLKMDIGFDWRLDFA
190 X 200 210

13. US-08-121-713B-33 (1-9)
R10694 Cephalosporin antibiotic biosynthetic enzyme #8.

ID R10694 standard; Protein; 437 AA.
AC R10694;
DT 27-MAR-1991 (first entry)
DE Cephalosporin antibiotic biosynthetic enzyme #8.
KW cephalosporin; antibiotic;
KW S-(L-alpha-aminoadipyl)-L-cysteine-D-; valine synthetase;
KW isopenicillin N synthetase; isopenicillin N epimerase;
KW deacetoxycephalosporin C synthetase; beta-lactamase;
KW deacetoxycephalosporin C hydroxylase.
OS Lysobacter lactamgenus.
PN J02291274-A.
PD 03-DEC-1990.
PF 10-JAN-1990; 003762.
PR 01-FEB-1989; JP-024710.
PR 10-JAN-1990; JP-003762.
PA (TAKE) TAKEDA CHEMICAL IND KK.

DR WPI; 91-018854/03.
DR N-PSDB; Q10190.
PT Prepn. of cephalosporin series antibiotics - comprises culturing
PT transformant of microbe transformed by plasmid contg. new DNA
PT fragment
PS Disclosure; Fig 20; 67pp; Japanese.
CC This protein is encoded by ORF8 of the 23666bp sequence
CC isolated from *L. lactamgenus* and comprising the genes for the
CC cephalosporin biosynthetic enzymes listed in the KEYWORDS. Plasmids
CC containing at least one of ORF's 1-9 can be used to transform
CC microbes, such as bacteria or yeast.
CC See also Q10191-2.
SQ Sequence 437 AA;
SQ 57 A; 30 R; 20 N; 18 D; 0 B; 4 C; 17 Q; 15 E; 0 Z; 31 G; 1 H;
SQ 23 I; 52 L; 8 K; 11 M; 19 F; 27 P; 24 S; 29 T; 11 W; 17 Y; 23 V;

Initial Score = 7 Optimized Score = 7 Significance = 3.45
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X X
WXXLXKXXL
PNERLRNMEWDLFKVPPESKISALLG
290 300 X 310

14. US-08-121-713B-33 (1-9)
R42078 Impatiens Necrotic Spot Virus S Non-structural pro
ID R42078 standard; Protein; 449 AA.
AC R42078;
DT 29-APR-1994 (first entry)
DE Impatiens Necrotic Spot Virus S Non-structural protein.
KW INSV; tospovirus; resistance; pan-handle secondary structure;
KW plant RNA virus.
OS Impatiens Necrotic Spot Virus.
PS EP-566525-A.
PD 20-OCT-1993.
PF 16-MAR-1993; 810190.
PR 19-MAR-1992; GB-006016.
PA (SANO) SANDOZ LTD.
PA (VGRJ/) VAN GRINSVEN M Q J M.
PA (SANO) SANDOZ PATENT GMBH.
PA (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.
PA (SANO) SANDOZ AG.

De Haan PT, Gielen JJJ, Goldbach RW, Peters D, Van Grinsven MQJ;
PI Van Grinsven MQJM;
PI WPI; 93-329814/42.
DR N-PSDB; Q49956.
PT Recombinant INSV DNA constructs comprising INSV DNA coding for
PT transcription into INSV RNA sequences - are used to transform
PT plants to reduce susceptibility to INSV infection
PS Example 4; Page 45-46; 64pp; English.
CC INSV RNA was purified from systemically infected *Nicotiana rustica*
CC leaves. The short (S) transcript comprises two genes, one coding for
CC a non-structural protein and the other (on the complementary strand)
CC coding for the nucleocapsid protein. The 5'- and 3'-terminal
CC sequences of the S RNA are capable of hybridising to each other; the
CC double-stranded structure obtained by such hybridisation is referred

CC to as a "pan-handle".
SQ Sequence 449 AA;
SQ 14 A; 14 R; 30 N; 23 D; 0 B; 5 C; 18 Q; 26 E; 0 Z; 15 G; 10 H;
SQ 30 I; 42 L; 35 K; 13 M; 22 F; 18 P; 55 S; 26 T; 5 W; 17 Y; 31 V;
Initial Score = 7 Optimized Score = 7 Significance = 3.45
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X X
WXXLXKXXL

VLSPTRSVHEWLYTLKPVFNQSQTNRTV
190 X 200 X 210

15. US-08-121-713B-33 (1-9)
R34135 C.roseus TDC with mutated N-terminus.

ID R34135 standard; Protein; 490 AA.
AC R34135;
DT 20-JUL-1993 (first entry)
DE C.roseus TDC with mutated N-terminus.
KW Tryptophan decarboxylase; tryptophan analogue; tdc gene.
OS Catharanthus roseus.
FH Key Location/Qualifiers
FT Protein 4..490
FT /label= TDC
FT /note= "mature N-terminus determined by direct
FT sequencing of isolated TDC"

FT Misc difference 2
FT /note= "Ser in this position in R34133"
FT Misc difference 10
FT /note= "Lys in this position in R34133"
PN W09306220-A.
PD 01-APR-1993.

PF 18-SEP-1992; E02175.
PR 20-SEP-1991; EP-202458.
PA (MOGE-) MOGEN INT NV.
PI Goddi Jn OJM, Hoge JHC, Schilperoort RA;
DR WPI; 93-117541/14.
DR N-PSDB; Q39283.
PT Selection method for transformed plants - comprises prodn. of
PT tryptophan analogues in plants and growth in medium deleterious
PT to non-transformed plant cells
PS Example 2; Page 32-34; 50pp; English.
CC An incomplete clone of C.roseus tdc gene was isolated. The missing
CC sequences were determined by primer extension on polyA+ RNA. (See
CC Q39281 for the determined nucleotide sequence). An oligonucleotide
CC was synthesised which encoded the missing bases spanning nucleotide
CC positions 90 to 125, flanked by a SalI and EcoRI restriction site
CC and containing an artificially introduced NcoI site. In the
CC synthetic sequence (see Q39282), there is a point mutation at
CC position 14 (corresponding to position 93 in Q39281) to create the
CC NcoI site. Due to an error in DNA synthesis also codon 10 (AAG in
CC Q39281) was changed to AAT. The resulting changes in encoded amino
CC acid sequence can be seen by comparing R34133 and R34135.
SQ Sequence 490 AA;
SQ 32 A; 23 R; 17 N; 28 D; 0 B; 9 C; 11 Q; 30 E; 0 Z; 26 G; 11 H;
SQ 29 I; 54 L; 26 K; 16 M; 21 F; 26 P; 36 S; 32 T; 10 W; 16 Y; 37 V;

Initial Score	=	7	Optimized Score	=	7	Significance	=	3.45
Residue Identity	=	33%	Matches	=	3	Mismatches	=	6
Gaps	=	0	Conservative Substitutions	=	0			

X
X
XXXXXX
X

FKNWQIATGRKFRSLKLWLIILRSYGVN
360 X 370 380

IntelliGenetics

FastDB -- Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq33pir.res made by maryh on Fri 19 May 95 10:45:29-PDT.

Query sequence being compared:US-08-121-713B-33 (1-9)
 Number of sequences searched: 75511
 Number of scores above cutoff: 3819

Results of the initial comparison of US-08-121-713B-33 (1-9) with:
Data bank : PIR 43, all entries

100000-
N
50000-
M
B
E
R
R
★
O
F10000-
S
E5000-
Q
U
E
N
C
E
S1000-
500-
100-
-

100000-
-
N
-
75000-
*

★

★

5000
1000
S E Q U E N C E S

505

→

100

[illegible]

PARAMETERS

	Similarity matrix	Unary	K-tuple
Mismatch penalty	1	Joining penalty	2
Gap penalty	1.00	Window size	20
Gap size penalty	0.05		6
Cutoff score	0		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2	5	1.66

Times:	CPU	Total Elapsed
	00:01:24.00	00:01:27.00

Number of residues:	22468834
Number of sequences searched:	75511
Number of scores above cutoff:	3819

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
1. S33400	Ig heavy chain - mouse	40	7	7	3.01 0
2. S39512	RNA polymerase beta', - red a	85	7	7	3.01 0
3. J01549	V1 protein - Panicum streak v	108	7	7	3.01 0

Brundish, H.; Markham, P.G.
J. Gen. Virol. (1992) 73:1041-1047
#journal The nucleotide sequence of an infectious insect-transmissible
#title clone of the geminivirus Panicum streak virus.
#contents Isolate Kenya
#accession JQ1549
#molecule_type DNA
#residues 1-108 ##label BRI
##cross-references EMBL:X60168
#length 108 #molecular-weight 11866 #checksum 2401
SUMMARY
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXXLKXXL
| |
FVSVLALYLLWLWIKDCICILLKAQRGS
50 60

4. US-08-121-713B-33 (1-9)
QOSABE hypothetical protein B-111 - Staphylococcus aureus
ENTRY QOSABE #type complete
TITLE hypothetical protein B-111 - Staphylococcus aureus plasmid
pe194
ORGANISM #formal name Staphylococcus aureus
DATE 18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change
30-Sep-1993
ACCESSION A04487
REFERENCE A91790
#authors Horinouchi, S.; Weisblum, B.
#journal J. Bacteriol. (1982) 150:804-814
#title Nucleotide sequence and functional map of pe194, a plasmid
that specifies inducible resistance to macrolide,
lincosamide, and streptogramin type B antibiotics.
#cross-references MUID:82167187
#accession A04487
##molecule_type DNA
##residues 1-111 ##label HOR

GENETICS
#genome plasmid
SUMMARY #length 111 #molecular-weight 13305 #checksum 8755
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXXLKXXL
| |

VLYPSAKAEWLEYIKELHIOFVVSPPLHD
30 X 40 50

5. US-08-121-713B-33 (1-9)

S32574 serum amyloid protein SAA - rabbit
ENTRY S32574 #type complete
TITLE serum amyloid protein SAA - rabbit
ORGANISM #formal name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 02-Dec-1993; #sequence_revision 02-Dec-1993; #text_change
02-Dec-1993

ACCESSIONS S32574
REFERENCE S32574
#authors Mitchell, T.I.; Coon, C.I.; Brinckerhoff, C.E.
#journal J. Clin. Invest. (1991) 87:1177-1185
#title Serum amyloid A (SAA) produced by rabbit synovial
fibroblasts treated with phorbol esters or interleukin 1
induces synthesis of collagenase and is neutralized with
specific antiserum.

#accession S32574
##status preliminary
##residues 1-122 ##label MIT
##cross-references EMBL:M64696
SUMMARY #length 122 #molecular-weight 13806 #checksum 1790
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXXLKXXL
| |
FLILGVNSREWLTLKEAGQGAKDMWRAY
20 30

6. US-08-121-713B-33 (1-9)
JN0029 GSAAL protein precursor - human
ENTRY JN0029 #type complete
TITLE GSAAL protein precursor - human
ORGANISM #formal name Homo sapiens #common_name man
DATE 10-Dec-1994; #sequence_revision 10-Dec-1994; #text_change
10-Dec-1994
ACCESSIONS JN0029
REFERENCE JN0029
#authors Sack Jr., G.H.; Talbot Jr., C.C.
#journal Gene (1989) 84:509-515
#title The human serum amyloid A (SAA)-encoding gene GSAAL:
nucleotide sequence and possible
autocrine-collagenase-inducer function.

#accession JN0029
##status preliminary
##residues 1-122 ##label SAC
SUMMARY #length 122 #molecular-weight 13440 #checksum 2835
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X

```
WXXXLKXXL
|  |
SIVLGVSSQGLTLFLKAAGOGAKDMWRAY
20 30
```

```
7. US-08-121-713B-33 (1-9)
   S23002      traJ protein - Escherichia coli plasmid RP4

ENTRY         S23002      #type complete
TITLE         traJ protein - Escherichia coli plasmid RP4
ORGANISM      #formal name Escherichia coli
DATE          31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change
              02-Aug-1994
ACCESSIONS    S23002
REFERENCE     S22992
AUTHORS       Ziegelin, G.; Pansegrau, W.; Strack, B.; Balzer, D.; Kroeger,
              M.; Kruit, V.; Lanka, E.
JOURNAL       DNA Seq. (1991) 1:303-327
TITLE         Nucleotide sequence and organization of genes flanking the
              transfer origin of promiscuous plasmid RP4.
CROSS-REFERENCES MUID:92190548
```

```
#accession S23002
#molecule type DNA
#residues 1-123 #label ZIE
#cross-references EMBL:X54459
#note the authors did not translate the codon for residue 1
```

```
GENETICS      traJ
#gene         plasmid
#genome       #length 123 #molecular-weight 13464 #checksum 2336
SEQUENCE
```

```
Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
```

```
X X
WXXXLKXXL
|  |
LARINGDLGRGLGLKLWLTDDPRTARFG
70 X 80 90
```

```
8. US-08-121-713B-33 (1-9)
   A34172      traJ protein - Escherichia coli plasmid RP4

ENTRY         A34172      #type complete
TITLE         traJ protein - Escherichia coli plasmid RP4
ORGANISM      #formal name Escherichia coli
DATE          07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change
              23-Mar-1993
ACCESSIONS    A34172
REFERENCE     A34172
AUTHORS       Ziegelin, G.; Fuerste, J.P.; Lanka, E.
JOURNAL       J. Biol. Chem. (1989) 264:11989-11994
TITLE         TraJ protein of plasmid RP4 binds to a 19-base pair invert
              sequence repetition within the transfer origin.
CROSS-REFERENCES MUID:89308605
#accession A34172
```

```
##molecule type DNA
##residues 1-123 #label ZIE
GENETICS      traJ
#gene         plasmid
#genome       #length 123 #molecular-weight 13464 #checksum 2336
SEQUENCE
```

```
Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
```

```
X X
WXXXLKXXL
|  |
LARINGDLGRGLGLKLWLTDDPRTARFG
70 X 80 90
```

```
9. US-08-121-713B-33 (1-9)
   S16481      hypothetical protein 1 - Sulfolobus solfataricus

ENTRY         S16481      #type complete
TITLE         hypothetical protein 1 - Sulfolobus solfataricus
ORGANISM      #formal name Sulfolobus solfataricus
DATE          03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
              03-Feb-1994
ACCESSIONS    S16481; S14845
REFERENCE     S16481
AUTHORS       Ramirez, C.; Matheson, A.T.
JOURNAL       Mol. Microbiol. (1991) 5:1687-1693
TITLE         A gene in the archaeobacterium Sulfolobus solfataricus that
              codes for a protein equivalent to the alpha subunits of the
              signal recognition particle receptor in eukaryotes.
```

```
#cross-references MUID:92048486
#accession S16481
#status preliminary
#molecule type DNA
#residues 1-126 #label RAM
#cross-references EMBL:X58538
#note the authors translated the codon GAA for residue 95 as
              Arg
```

```
SUMMARY      #length 126 #molecular-weight 14744 #checksum 8684
SEQUENCE
```

```
Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
```

```
X X
WXXXLKXXL
|  |
RKFWMTIESNSWLKYLKIYKVNYSLSQ
110 X 120
```

```
10. US-08-121-713B-33 (1-9)
   S22996      traJ protein - Escherichia coli plasmid R751

ENTRY         S22996      #type complete
```

```
TITLE      traJ protein - Escherichia coli plasmid R751
ORGANISM   #formal name Escherichia coli
DATE       31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change
          02-Aug-1994
ACCESSIONS S22996
REFERENCE   Ziegelin, G.; Pansegrau, W.; Strack, B.; Balzer, D.; Kroeger,
          M.; Kruft, V.; Lanka, E.
#journal   DNA Seq. (1991) 1:303-327
#title     Nucleotide sequence and organization of genes flanking the
          transfer origin of promiscuous plasmid RP4.
#cross-references MUID:92190548
#accession  S22996
##molecule_type DNA
##residues 1-130 #label ZIE
##cross-references EMBL:X54458
GENETICS
#gene      traJ
#genome    plasmid
SUMMARY    #length 130 #molecular-weight 14392 #checksum 9839
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X      X
WXXXLKXXL
|||
IVRVNGDLGRGLGLKLLWLTDDVDTLQFG
70      80      90

11. US-08-121-713B-33 (1-9)
    JV0111 probable transcriptional activator fecI - Escheric

ENTRY      JV0111 #type complete
TITLE      probable transcriptional activator fecI - Escherichia coli
ORGANISM   #formal name Escherichia coli
DATE       12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
          18-Jun-1993
ACCESSIONS JV0111
REFERENCE   JV0111
#authors   Van Hove, B.; Staudenmaier, H.; Braun, V.
#journal   J. Bacteriol. (1990) 172:6749-6758
#title     Novel two-component transmembrane transcription control:
          regulation of iron dicitrate transport in Escherichia coli
          K-12.
#cross-references MUID:91072220
#contents  Strain K12
#accession JV0111
##molecule_type DNA
##residues 1-173 #label VAN
GENETICS
#gene      fecI
#map_position 93 min
KEYWORDS   membrane protein; transcription activator
SUMMARY    #length 173 #molecular-weight 19480 #checksum 8614
SEQUENCE
```

```
Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X      X
WXXXLKXXL
|||
SLTFESLYGTHHGWLKSWLTKRLQSAFDA
10      20      30

12. US-08-121-713B-33 (1-9)
    S44695 hypothetical membrane protein YBR116c - yeast (Saccharomyces
ENTRY      S44695 #type complete
TITLE      hypothetical membrane protein YBR116c - yeast (Saccharomyces
          cerevisiae)
ALTERNATE_NAMES hypothetical protein YBR0911
ORGANISM   #formal name Saccharomyces cerevisiae
DATE       27-Jun-1994 #sequence_revision 09-Sep-1994 #text_change
          09-Sep-1994
ACCESSIONS S44695; S45984
REFERENCE   S44670
#authors   Feldmann, H.
#submission submitted to the EMBL Data Library, April 1994
#accession S44695
##molecule_type DNA
##residues 1-175 #label FEL
##cross-references EMBL:X78993
REFERENCE   S45927
#authors   Feldmann, H.; Mannhaupt, G.; Schwarziolse, C.; Vetter, I.
#submission submitted to the Protein Sequence Database, August 1994
#accession S45984
##molecule_type DNA
##residues 1-175 #label FE2
##cross-references EMBL:Z35985
GENETICS
#map_position 2R
KEYWORDS   membrane protein
FEATURE    120-136 #domain transmembrane #status predicted #label TM1\
          138-158 #domain transmembrane #status predicted #label TM2
SUMMARY    #length 175 #molecular-weight 20215 #checksum 3013
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X      X
WXXXLKXXL
|||
TDSLIFYQTVFRSCPILKYLLQAGVSMLIN
40 X      50      60

13. US-08-121-713B-33 (1-9)
    A39414 flavodoxin - Enterobacter agglomerans plasmid pEA3
ENTRY      A39414 #type complete
```

TITLE flavodoxin - Enterobacter agglomerans plasmid pEA3
ALTERNATE_NAMES electron transport protein nifF
ORGANISM #formal name Enterobacter agglomerans
DATE 21-Feb-1992 #sequence_revision 21-Feb-1992 #text_change
20-Aug-1994
ACCESSIONS A39414
REFERENCE A39414
#authors Kreutzer, R.; Dayananda, S.; Klingmuller, W.
#journal J. Bacteriol. (1991) 173:3252-3256
#title Co-transcription of the electron transport protein genes nifF
and nifH in Enterobacter agglomerans 333.
#cross-references MUID:91217003
#accession A39414
##molecule_type DNA
##residues 1-177 ##label KRE
##cross-references GB:M38221
GENETICS
#gene nifF
#genome plasmid
CLASSIFICATION #superfamily flavodoxin; flavodoxin homology
KEYWORDS electron transfer; flavoprotein; FMN
SUMMARY #length 177 #molecular-weight 19581 #checksum 1188
SEQUENCE
Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
X X
WXXLKKXXL
| |
YDLTEERIDSWLEKLPKAVL
160 170 X

14. US-08-121-713B-33 (1-9)
RBY36 cell division control protein CDC36 - yeast (Sacch
cerevisiae)
ENTRY RBY36 #type complete
TITLE cell division control protein CDC36 - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES start protein CDC36
ORGANISM #formal name Saccharomyces cerevisiae
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
05-May-1994
ACCESSIONS S12304; A26372
REFERENCE S05830
#authors Barker, D.G.; White, J.H.M.; Johnston, L.H.
#journal Nucleic Acids Res. (1985) 13:8323-8337
#title The nucleotide sequence of the DNA ligase gene (CDC9) from
Saccharomyces cerevisiae: a gene which is cell-cycle
regulated and induced in response to DNA damage.
#cross-references MUID:86093646
#accession S12304
##molecule_type DNA
##residues 1-191 ##label BAR
##cross-references EMBL:X03246
REFERENCE A93635
#authors Ferguson, J.; Ho, J.Y.; Peterson, T.A.; Reed, S.I.
#journal Nucleic Acids Res. (1986) 14:6681-6697

#title Nucleotide sequence of the yeast cell division cycle start
genes CDC28, CDC36, CDC37, and CDC39, and a structural
analysis of the predicted products.
#cross-references MUID:86312926
#accession A26372
##molecule_type DNA
##residues 1-191 ##label FER
##cross-references GB:X04287
GENETICS
#gene LISTA:CDC36
#map_position 4L
CLASSIFICATION #superfamily cell division control protein CDC36
KEYWORDS cell cycle control
SUMMARY #length 191 #molecular-weight 22362 #checksum 7370
SEQUENCE
Initial Score = 7 Optimized Score = 7 Significance = 3.01
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0
X X
WXXLKKXXL
| |
YELRKRWYHKTLLKAWLTQPMMEPIV
130 140 150
15. US-08-121-713B-33 (1-9)
PCL136 purF protein - Lactobacillus casei (fragment)
ENTRY PCL136 #type fragment
TITLE purF protein - Lactobacillus casei (fragment)
ORGANISM #formal name Lactobacillus casei
DATE 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change
14-Sep-1994
ACCESSIONS PCL136; PC2004
REFERENCE JC1290
#authors Gu, Z.M.; Martindale, D.W.; Lee, B.H.
#journal Gene (1992) 119:123-126
#title Isolation and complete sequence of the purL gene encoding
FGM synthase II in Lactobacillus casei.
#cross-references MUID:93012962
#accession PCL136
##molecule_type DNA
##residues 1-5, 'RLKMKNAGFRC', 18-194 ##label GUZ
REFERENCE JC2024
#authors Gu, Z.M.; Martindale, D.W.; Lee, B.H.
#journal Gene (1993) 133:147
#title Corrigendum: Isolation and complete sequence of the purL gene
encoding FGM synthase II in Lactobacillus casei.
#accession PC2004
##molecule_type DNA
##residues 1-19 ##label GU2
##note this sequence corrects amino terminal sequence errors of
purF protein in PCL136
SUMMARY #length 194 #checksum 5649
SEQUENCE
Initial Score = 7 Optimized Score = 9 Significance = 3.01
Residue Identity = 44% Matches = 4 Mismatches = 5

Gaps = 0 Conservative Substitutions = 0

X X
WXXLXXL
| | |
HLIRRVGQPWLTLKLTALNEVHGGAFFV
150 X 160 X 170

IntelliGenetics

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq33spt.res made by on Fri 19 May 95 9:01:55-PDT.

Query sequence being compared:US-08-121-713B-33 (1-9)
Number of sequences searched: 43470
Number of scores above cutoff: 4864

Results of the initial comparison of US-08-121-713B-33 (1-9) with:
Data bank : Swiss-Prot 31, all entries

100000-
-
-
N
U5000-
-
M
B
E
E
R
-
-
O
F10000-
-
-
*

Age Group	Number of Respondents
18-24	480
25-34	420
35-44	380
45-54	320
55-64	280
65-74	220
75-84	180
85-94	120
95+	10

Rating	Number of stars
1	100
2	150
3	250
4	350
5	450

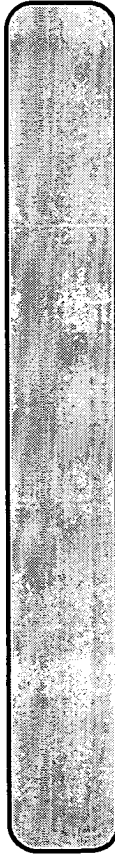
✱

A 100% similar sequence to the query sequence was found:

Sequence Name	Description	Length	Score	Init. Opt.
1. PLSX_ECOLI	PLSX PROTEIN.	346	9	4.04

The list of other best scores is:

The list of other best scores is:



PARAMETERS

Similarity matrix	Unitary	K-tuple
Mismatch penalty	1	Joining penalty
Gap penalty	1.00	Window size
Gap size penalty	0.05	
Cutoff score	0	
Randomization group	0	

Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores: Mean 3 Median 5 Standard Deviation 1.49

Times:	CPU	Total Elapsed
	00:00:45.00	00:00:45.00

Number of residues:	15335248
Number of sequences searched:	43470
Number of scores above cutoff:	4864

Cut-off raised to 2.
Cut-off raised to 4.
Cut-off raised to 5.

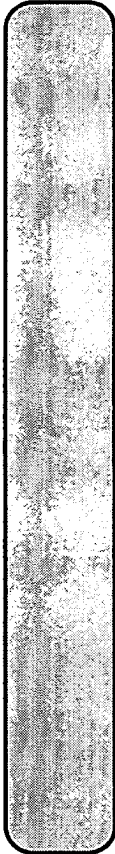
The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% similar sequence to the query sequence was found:

Sequence Name	Description	Length	Score	Init. Opt.
1. PLSX_ECOLI	PLSX PROTEIN.	346	9	4.04

The list of other best scores is:

The list of other best scores is:



Sequence Name	Description	Length	Init. Opt.	Score	Sig.	Frame
**** 2 standard deviations above mean ****						
2. ARRE_RAT	ARRESTIN-E (FRAGMENT).	72	7	2.69	0	
3. RPOD_GALSU	DNA-DIRECTED RNA POLYMERASE B	85	7	2.69	0	
4. Y1IK_PASVK	HYPOTHETICAL 11.9 KD PROTEIN	108	7	2.69	0	
5. TRJ6_ECOLI	TRAJ PROTEIN (RELAXOSOME PROT	122	7	2.69	0	
6. SAA3_RABIT	RABBIT SERUM AMYLOID A (SAA3)	122	7	2.69	0	
7. SAA3_HUMAN	SERUM AMYLOID A-3 PROTEIN PRE	122	7	2.69	0	
8. YDP1_SULSO	HYPOTHETICAL 14.7 KD PROTEIN	126	7	2.69	0	
9. CYT_GYPCA	CYSTATIN PRECURSOR.	129	7	2.69	0	
10. TRJ5_ECOLI	TRAJ PROTEIN.	137	7	2.69	0	
11. YZ21_ECOLI	HYPOTHETICAL 15.2 KD PROTEIN	137	7	2.69	0	
12. YR65_SCHPO	HYPOTHETICAL 15.8 KD PROTEIN	137	7	2.69	0	
13. FEG1_ECOLI	PROBABLE RNA POLYMERASE SIGMA	173	7	2.69	0	
14. YBW6_YEAST	VERY HYPOTHETICAL 20.2 KD PRO	175	7	2.69	0	
15. FLAV_ENTAG	FLAVODOXIN.	177	7	2.69	0	
16. NOT2_YEAST	GENERAL NEGATIVE REGULATOR OF	191	7	2.69	0	
17. PUR1_LACCA	AMIDOPHOSPHORIBOSYLTRANSFERAS	193	7	2.69	0	
18. REPY_STAAD	REPLICATION PROTEIN.	199	7	2.69	0	
19. FIMB_ECOLI	TYPE 1 FIMBRIAE REGULATORY PR	200	7	2.69	0	
20. YS2 SYN7	HYPOTHETICAL 23.0 KD PROTEIN	208	7	2.69	0	
21. ARA1_ARATH	RAS-RELATED PROTEIN ARA-1.	218	7	2.69	0	
22. PIGF_HUMAN	PHOSPHATIDYLINOSITOL-GLYCAN B	219	7	2.69	0	
23. PRT1_HANPO	POTATIVE PRT1 PROTEIN.	220	7	2.69	0	
24. ATP6_BACCA	ATP SYNTHASE A CHAIN (EC 3.6.	237	7	2.69	0	
25. RM08_YEAST	MITOCHONDRIAL 60S RIBOSOMAL P	238	7	2.69	0	
26. ATP6_THERP3	ATP SYNTHASE A CHAIN (EC 3.6.	238	7	2.69	0	
27. ARTQ_ECOLI	ARGININE TRANSPORT SYSTEM PER	238	7	2.69	0	
28. FC8B_HUMAN	HIGH AFFINITY IMMUNOGLOBULIN	244	7	2.69	0	
29. YN70_YEAST	HYPOTHETICAL 27.5 KD PROTEIN	246	7	2.69	0	
30. TRFC_SULSO	INDOLE-3-GLYCEROL PHOSPHATE S	248	7	2.69	0	
31. YAFD_ECOLI	HYPOTHETICAL 29.1 KD PROTEIN	259	7	2.69	0	
32. TERM_BPPRD	DNA TERMINAL PROTEIN (PROTEIN	259	7	2.69	0	
33. YPOQ_KLEPN	HYPOTHETICAL PROTEIN IN PQQA	271	7	2.69	0	
34. ARGB_PORUM	ACETYLGLOUTAMATE KINASE (EC 2.	283	7	2.69	0	
35. YLJE_ECOLI	HYPOTHETICAL 34.1 KD PROTEIN	312	7	2.69	0	
36. A26L_HUMAN	LEUCINE-RICH ALPHA-2-GLYCOPRO	312	7	2.69	0	
37. Y085_CAEEL	HYPOTHETICAL 35.7 KD PROTEIN	313	7	2.69	0	
38. REPA_NEIGO	REPLICATION PROTEIN.	328	7	2.69	0	
39. FBP_NEIME	MAJOR FERRIC IRON BINDING PRO	330	7	2.69	0	
40. FBP_NEIGO	MAJOR FERRIC IRON BINDING PRO	330	7	2.69	0	
41. COX2_THERP3	CYTOCHROME C OXIDASE POLYPEPT	356	7	2.69	0	
42. AR0G_BACSU	PHOSPHO-2-DEHYDRO-3-DEOXYTHEP	358	7	2.69	0	
43. WNT2_MOUSE	WNT-2 PROTEIN PRECURSOR (IRP	360	7	2.69	0	
44. WNT2_HUMAN	WNT-2 PROTEIN PRECURSOR (IRP	360	7	2.69	0	
45. CD14_MOUSE	MONOCYTE DIFFERENTIATION ANTI	366	7	2.69	0	

1. US-08-121-713B-33 (1-9)

PLSX_ECOLI PLSX PROTEIN.

ID PLSX_ECOLI STANDARD; PRT; 346 AA.

AC P272A7;

DT 01-AUG-1992 (REL. 23, CREATED)

DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)

DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)

DE PLSX PROTEIN.

GN PLSX.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILLICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RM 93077479
RA OH W., LARSON T.J.;
RL J. BACTERIOL. 174:7873-7874(1992).
CC -!- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN FATTY ACID OR
CC PHOSPHOLIPID SYNTHESIS.
DR EMBL; M96793; ECPLSPABA.
DR ECOGENE; EG11437; PLSX.
KW FATTY ACID BIOSYNTHESIS; PHOSPHOLIPID BIOSYNTHESIS.
SQ SEQUENCE 346 AA; 37100 MW; 592849 CN;
Initial Score = 9 Optimized Score = 9 Significance = 4.04
Residue Identity = 44% Matches = 4 Mismatches = 5
Gaps = 0 Conservative Substitutions = 0

X X
WXXLXXL
- - -
SQEGKGRSWLLLLKRWLQKSLTRFSH
250 X 260 X 270

2. US-08-121-713B-33 (1-9)
ARRE_RAT ARRESTIN-E (FRAGMENT).

ID ARRE_RAT STANDARD; PRT; 72 AA.

AC P3720;

DT 01-OCT-1994 (REL. 30, CREATED)

DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)

DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)

DE ARRESTIN-E (FRAGMENT).

GN EAR.

OS RATTUS NORVEGICUS (RAT).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; RODENTIA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY; TISSUE=PINEAL GLAND;

RM 94140898

RA CRAFT C.M., WHITMORE D.H., WIECHMANN A.F.;

RL J. BIOL. CHEM. 269:4613-4619(1994).

CC -!- TISSUE SPECIFICITY: ADRENAL, CEREBRAL CORTEX, HEART, HYPOTHALAMUS,
CC INTESITINE, LIVER, LUNG, PITUITARY, RETINA, TESTIS.CC -!- SIMILARITY: SIMILARITY IN THE C-TERMINAL WITH ALPHA-TRANSDUCIN
CC AND OTHER PURINE NUCLEOTIDE-BINDING PROTEINS.

CC -!- SIMILARITY: BELONGS TO THE ARRESTIN FAMILY.

DR EMBL; U03630; R003630.

KW SENSORY TRANSDUCTION.

FT NON TER 1

SQ SEQUENCE 72 AA; 8132 MW; 28630 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

DTNLHSSQOCWLCIKNSKHNVISSRGG
10 X 20 X 30
X X
WXXLKKXXL
|
|

3. US-08-121-713B-33 (1-9)
RPOD_GALSU DNA-DIRECTED RNA POLYMERASE BETA* CHAIN (EC 2.7.7.7).

ID RPOD_GALSU STANDARD; PRT; 85 AA.
AC P35018;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE DNA-DIRECTED RNA POLYMERASE BETA* CHAIN (EC 2.7.7.6) (FRAGMENT).
GN RPOC2.
OS GARDIERIA SULPHURARIA (CYANIDIUM CALDARIUM).
OC CHLOROPLAST.
OC EUKARYOTA; PLANTA; PHYCOPHYTA; RHODOPHYTA (RED ALGAE).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=14-1-1 / ISOLATE 107.79/GOETTINGEN;
RM 94033298
RA KOSTRZEWA M., ZETSCHKE K.;
RL PLANT MOL. BIOL. 23:67-76(1993).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR
CC SUBUNITS: ALPHA, BETA, BETA', AND BETA".
DR EMBL; X67814; GSPLSEQ.
DR PIR; S39512; S39512.
KW TRANSCRIPTION; DNA-DIRECTED RNA POLYMERASE; CHLOROPLAST.
FT NON TER
SQ SEQUENCE 85 AA; 10114 MW; 34876 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXLKKXXL
|
|
TKAUFENHIDWLKGLKENVIIIGRLIPACT
10 20 30

4. US-08-121-713B-33 (1-9)
Y11K_PASVK HYPOTHETICAL 11.9 KD PROTEIN (ORF VI).

ID Y11K_PASVK STANDARD; PRT; 108 AA.
AC Q00336;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 11.9 KD PROTEIN (ORF VI).

OS PANICUM STREAK VIRUS (KENYAN ISOLATE).
OC VIRIDAE; SS-DNA NONENVELOPED VIRUSES; GEMINIVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RM 92268861
RA BRIDDON R.W., LUNNESS P., CHAMBERLAIN L.C., BRUNDISH H.,
RA PINNER M.S., MARKHAM P.G.;
RL J. GEN. VIROL. 73:1041-1047(1992).
DR EMBL; X60168; PSGIITDNA.
DR PIR; J01549; J01549.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 108 AA; 11866 MW; 63187 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXLKKXXL
|
|
FVSVLALYLILWLWIKDKCILLIKQAQRGRS
50 60

5. US-08-121-713B-33 (1-9)
TRJ6_ECOLI TRAJ PROTEIN (RELAXOSOME PROTEIN).

ID TRJ6_ECOLI STANDARD; PRT; 122 AA.
AC P17909;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE TRAJ PROTEIN (RELAXOSOME PROTEIN).
GN TRAJ.
OS ESCHERICHIA COLI.
OG PLASMID INCP-BETA RP4.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-5 AND 118-122.
RM 89308605
RA ZIEGELIN G., FUERSTE J.P., LANKA E.;
RL J. BIOL. CHEM. 264:11989-11994(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HB101;
RM 92190548
RA ZIEGELIN G., PANSEGRAU W., STRACK B., BALZER D., KROEGER M., KROFT V.,
RA LANKA E.;
RL DNA SEQ. 1:303-327(1991).
RN [3]
RP SEQUENCE OF 1-7 FROM N.A.
RM 89184510
RA LANKA E., FUERSTE J.P.;
RL PROC. NATL. ACAD. SCI. U.S.A. 86:1771-1775(1989).
CC -!- FUNCTION: TRANSFER OF PLASMID RP4 DURING BACTERIAL CONJUGATION
CC REQUIRES THE PLASMID-ENCODED TRAJ PROTEIN, WHICH BINDS TO A 19-
CC BASE PAIR INVERT SEQUENCE REPETITION WITHIN THE TRANSFER ORIGIN.
CC TRAJ PROTEIN IS BOUND TO ONLY ONE SIDE OF THE DNA HELIX. THIS
CC NUCLEOPROTEIN STRUCTURE IS THE INITIAL COMPLEX IN THE PATHWAY TO

CC ASSEMBLE A FUNCTIONAL RELAXOSOME.

CC -I- SUBUNIT: MONOMER.

DR EMBL; J04942; PRTRAJ.

DR EMBL; X54459; ECRP4.

DR EMBL; M25423; PRTRAB.

DR PIR; A34172; A34172.

DR PIR; S23002; S23002.

KW CONJUGATION; DNA-BINDING; PLASMID.

FT INIT MET 0 0

SQ SEQUENCE 122 AA; 13332 MW; 70495 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXLKXXL

| |

LARINGDLGRIGGLIKWLTDPTARFG

70 X 80

6. US-08-121-713B-33 (1-9)

SAA3 RABBIT RABBIT SERUM AMYLOID A (SAA3) MRNA, COMPLETE CDS.

ID SAA3 RABBIT STANDARD; PRT; 122 AA.

AC P35573;

DT 01-JUN-1994 (REL. 29, CREATED)

DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)

DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)

DE RABBIT SERUM AMYLOID A (SAA3) MRNA, COMPLETE CDS.

GN SAA3.

OS ORYCTOLAGUS CUNICULUS (RABBIT).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; LAGOMORPHA.

RN [1]

RP SEQUENCE FROM N.A.

RM 91185595

RA MITCHELL T.I.; COON C.I.; BRINCKERHOFF C.E.;

RL J. CLIN. INVEST. 87:1177-1185(1991).

CC -I- FUNCTION: MAJOR ACUTE PHASE REACTANT. APOLIPOPROTEIN OF THE HDL

CC COMPLEX.

CC -I- INDUCTION: UPON CYTOKINE STIMULATION.

CC -I- TISSUE SPECIFICITY: PLASMA, SYNTHESIZED BY THE LIVER.

CC -I- DISEASE: REACTIVE. SECONDARY AMYLOIDOSIS IS CHARACTERIZED BY THE

CC EXTRACELLULAR ACCUMULATION IN VARIOUS TISSUES OF THE SAA PROTEIN.

CC THESE DEPOSITS ARE HIGHLY INSOLUBLE AND RESISTANT TO PROTEOLYSIS;

CC THEY DISRUPT TISSUE STRUCTURE AND COMPROMISE FUNCTION.

CC -I- SIMILARITY: TO OTHER MAMMALIAN SAAS.

CC EMBL; M64696; OCSAA3.

DR PIR; S32574; S32574.

DR PROSITE; PS00992; SAA.

KW ACUTE PHASE; PLASMA; HDL; AMYLOID; SIGNAL; MULTIGENE FAMILY.

FT SIGNAL 1 18 BY SIMILARITY.

FT CHAIN 19 122 SERUM AMYLOID A-3 PROTEIN.

SQ SEQUENCE 122 AA; 13806 MW; 64764 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXLKXXL

| |

FLILGVNSREWLTFLEAGQGAKDMWRAY

20 30

7. US-08-121-713B-33 (1-9)
SAA3_HUMAN SERUM AMYLOID A-3 PROTEIN PRECURSOR.

ID SAA3 HUMAN STANDARD; PRT; 122 AA.

AC P22614;

DT 01-AUG-1991 (REL. 19, CREATED)

DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)

DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)

DE SERUM AMYLOID A-3 PROTEIN PRECURSOR.

GN SAA3.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RM 90128298

RA SACK G.H. JR.; TALBOT C.C. JR.;

RL GENE 84:509-515(1989).

CC -I- FUNCTION: PROBABLY A PSEUDOGENE.

CC -I- DISEASE: REACTIVE. SECONDARY AMYLOIDOSIS IS CHARACTERIZED BY THE

CC EXTRACELLULAR ACCUMULATION IN VARIOUS TISSUES OF THE SAA PROTEIN.

CC THESE DEPOSITS ARE HIGHLY INSOLUBLE AND RESISTANT TO PROTEOLYSIS;

CC THEY DISRUPT TISSUE STRUCTURE AND COMPROMISE FUNCTION.

CC -I- SIMILARITY: TO OTHER MAMMALIAN SAAS.

DR EMBL; X13895; HSSAALA.

DR PIR; JN0029; JN0029.

DR PROSITE; PS00992; SAA.

KW ACUTE PHASE; PLASMA; HDL; AMYLOID; SIGNAL; MULTIGENE FAMILY.

FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 19 122 SERUM AMYLOID A-3 PROTEIN.

SQ SEQUENCE 122 AA; 13440 MW; 70025 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
XXXXLKXXL

| |

SILVLGVSSQGLWTLFLKAAAGQGAKDMWRAY

20 30

8. US-08-121-713B-33 (1-9)
YDPI_SULSO HYPOTHETICAL 14.7 KD PROTEIN IN DPA 5'REGION (ORF

ID YDPI SULSO STANDARD; PRT; 126 AA.

AC P27339;

DT 01-AUG-1992 (REL. 23, CREATED)

DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)

DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 14.7 KD PROTEIN IN DPA 5'REGION (ORF 1) (A03).

OS SULFOLOBUS SOLFATARICUS.
OC PROKARYOTA; MENDOSICUTES; ARCHAEABACTERIA; SULFOLOBALES.

[1]
PP SEQUENCE FROM N.A.
RC STRAIN=DSM 1616 / P1;
RM 92048486
RA RAMIREZ C., MATHESON A.T.;
RL MOL. MICROBIOL. 5:1687-1693(1991).
DR EMBL; X58538; SSDOCK.
DR PIR; S14845; S14845.
DR PIR; S16481; S16481.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 126 AA; 14744 MW; 87526 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X X
WXXLKKXXL
||
RKFWIESNSWLKYLKYNLSQ
110 X 120

9. US-08-121-713B-33 (1-9)

CYT_CYPCA CYSTATIN PRECURSOR.

ID CYT_CYPCA STANDARD; PRT; 129 AA.
AC P35481;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE CYSTATIN PRECURSOR.

OS CYPRINUS CARPIO (COMMON CARP).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;
OC OSTEICHTHYES; ACTINOPTERYGII; CYPRINIFORMES.

[1]
PP SEQUENCE FROM N.A.
RC TISSUE=OVARY;
RA TSAI Y., HUANG F.L.;
RL SUBMITTED (XXX-1993) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- FUNCTION: CYSTEINE PROTEINASE INHIBITOR.
CC -1- SIMILARITY: THIS IS A TYPE 2 CYSTATIN.

DR EMBL; L23572; CCGYST.
DR HSP; P01038; ICEW.
DR THIOLESTERASE INHIBITOR; SIGNAL.

FT SIGNAL 1 ? POTENTIAL.
FT CHAIN ? 129 CYSTATIN.
FT ACT_SITE 23 23 REACTIVE SITE (BY SIMILARITY).
FT SITE 67 71 SECONDARY AREA OF CONTACT.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 108 128 BY SIMILARITY.
SQ SEQUENCE 129 AA; 14236 MW; 95323 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXLKKXXL
||
OCKITVWSOPWLSIKVTENTCM
110 120 X

10. US-08-121-713B-33 (1-9)

TRJ5_ECOLI TRAJ PROTEIN.

ID TRJ5_ECOLI STANDARD; PRT; 130 AA.
AC P17907;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE TRAJ PROTEIN.
GN TRAJ.

OS ESCHERICHIA COLI.

OG PLASMID INCP-BETA R751.

OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;

OC ENTEROBACTERIACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HB101;

RM 92190548

RA ZIEGELIN G., PANSEGRAU W., STRACK B., BALZER D., KROEGER M.,

RA KRUFFT V., LANKA E.;

RL DNA SEQ. 1:303-327(1991).

RN [2]

RP SEQUENCE OF 1-16 FROM N.A.

RM 89184510

RA LANKA E., FUERSTE J.P.;

RL PROC. NATL. ACAD. SCI. U.S.A. 86:1771-1775(1989).

CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR POSITIVELY REGULATING THE
CC EXPRESSION OF DNA BETWEEN BACTERIAL CELLS.
CC TRANSFER OF DNA BETWEEN BACTERIAL CELLS.

CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.

DR EMBL; X54458; ECR751.

DR EMBL; M25422; PRITRA.

DR PIR; S22996; S22996.

KW OUTER MEMBRANE; CONJUGATION; TRANSCRIPTION REGULATION; ACTIVATOR;

KW PLASMID.

SQ SEQUENCE 130 AA; 14392 MW; 79743 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXLKKXXL
||

LVRVNGDLGRGLGLIKLWLTDDVRLQFG
70 80 90

11. US-08-121-713B-33 (1-9)

YZZI_ECOLI HYPOTHETICAL 15.2 KD PROTEIN IN ICC 3'REGION.

ID YZZI_ECOLI STANDARD; PRT; 137 AA.
AC P36653;

DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DE 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 15.2 KD PROTEIN IN ICC 3' REGION.
GN YZZI.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA IMAMURA K., NIKI H., YAMANAKA K., OGURA T., FUJITA N., ISHIHAMA A.,
RA HIRAGA S.;
RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; D16557; ECICC.
DR ECOGENE; EGI2186; YZZI.
KW HYPOTHETICAL PROTEIN;
SQ SEQUENCE 137 AA; 15256 MW; 102300 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X X
WXXXLKXXL
||
HGNSPRSASILLKNWLAHHHPDVEMI
10 20 X 30

12. US-08-121-713B-33 (1-9)
YR65_SCHPO HYPOTHETICAL 15.8 KD PROTEIN IN REC6 5' REGION (ORF

ID YR65_SCHPO STANDARD; PRT; 137 AA.
AC P40922;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 15.8 KD PROTEIN IN REC6 5' REGION (ORF137).
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RM 94274029
RA LIN Y., SMITH G.R.;
RL GENETICS 136:769-779(1994).
DR EMBL; L14773; SPREC6A.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 137 AA; 15776 MW; 107988 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X X
WXXXLKXXL
||
KPIKLSLSNSTPLLLKILMLIVAVTWVVLV
10 20 X 30

13. US-08-121-713B-33 (1-9)
FECI_ECOLI PROBABLE RNA POLYMERASE SIGMA FACTOR FECI.

ID FECI_ECOLI STANDARD; PRT; 173 AA.
AC P23484;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE PROBABLE RNA POLYMERASE SIGMA FACTOR FECI.
GN FECI.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RM 91072220
RA VAN HOVE B., STAUDENMAIER H., BRAUN V.;
RL J. BACTERIOL. 172:6749-6758(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MGI655;
RA BURLAND V.D., PLONKETT G. III, BLATTNER F.R.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SIMILARITY TO OTHER ECF SIGMA FACTORS.
RM 94329558

RA LONETTO M.A., BROWN K.L., RUDD K.E., BUTTNER M.J.;
RL PROC. NATL. ACAD. SCI. U.S.A. 91:7573-7577(1994).
CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
THEN IS RELEASED. THIS SIGMA FACTOR REGULATES THE FEC GENES FOR
IRON DICITRATE TRANSPORT (PROBABLE).
CC -!- SIMILARITY: TO OTHER SIGMA FACTORS THAT DO NOT BELONG TO THE
SIGMA-54 FAMILY. BELONGS TO THE ECF SUBFAMILY.
DR EMBL; M63115; ECFECIR.
DR EMBL; U14003; ECUW93.
DR PIR; JVO111; JVO111.
DR ECOGENE; EGI0291; FECI.
KW TRANSCRIPTION REGULATION; SIGMA FACTOR; DNA-DIRECTED RNA POLYMERASE;
KW DNA-BINDING; IRON TRANSPORT; TRANSPORT.
FT DOMAIN 40 52 H-T-H MOTIF (BY SIMILARITY).
FT DNA BIND 139 158
SQ SEQUENCE 173 AA; 19480 MW; 155397 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X X
WXXXLKXXL
||
SLTFESLYGTHHGWLKSWLTRKLSAFDA
10 20 30

14. US-08-121-713B-33 (1-9)
YBW6_YEAST VERY HYPOTHETICAL 20.2 KD PROTEIN IN LYS2-TKL2 INT

ID YW6 YEAST STANDARD; PRT; 175 AA.
AC P38268;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE VERY HYPOTHETICAL 20.2 KD PROTEIN IN LYS2-TKL2 INTERGENIC REGION.
GN YBR116C OR YBR0911.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA MANNHAUPT G., STUCKA R., EHMLE S., VETTER I., FELDMANN H.;
RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X78993; SCRACII.
DR EMBL; Z35985; SCYBR116C.
DR PIR; S44695; S44695.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 143 166 POTENTIAL.
SQ SEQUENCE 175 AA; 20215 MW; 175283 CN;

Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXXLKXXL
|||
TDSLFTVFRSCPLKYWLLQAGVSMLIN
40 X 50 60

15. US-08-121-713B-33 (1-9)
FLAV_ENTAG FLAVODOXIN.

ID FLAV_ENTAG STANDARD; PRT; 177 AA.
AC P28579;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE FLAVODOXIN.
GN NIFF.
OS ENTEROBACTER AGGLOMERANS.
OC PLASMID PEAS.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=333;
RM 91217003
RA KREUTZER R., DAYANANDA S., KLINGMUELLER W.;
RL J. BACTERIOL. 173:3252-3256(1991).
CC -1- FUNCTION: ELECTRON-TRANSFER PROTEINS THAT FUNCTION IN VARIOUS
CC ELECTRON TRANSPORT SYSTEMS IN MICRO-ORGANISMS. FUNCTIONALLY
CC INTERCHANGEABLE WITH FERREDOXIN.
CC -1- COFACTOR: FMN.
DR EMBL; M38221; PPNIFF.
DR PIR; A39414; A39414.
DR HSP; P10340; 10FV.
DR PROSITE; PS00201; FLAVODOXIN.

KW NITROGEN FIXATION; ELECTRON TRANSPORT; FLAVOPROTEIN; FMN; PLASMID.
SQ SEQUENCE 177 AA; 19581 MW; 159879 CN;
Initial Score = 7 Optimized Score = 7 Significance = 2.69
Residue Identity = 33% Matches = 3 Mismatches = 6
Gaps = 0 Conservative Substitutions = 0

X X
WXXXLKXXL
|||
YDLTEERIDSWLEKLPKPAVL
160 170 X

maryh@stic

stdin

NeWSprinter20

Fri May 19 11:07:02 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5

IntelliGenetics

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq43asq.res made by on Fri 19 May 95 8:43:41-PDT.

Query sequence being compared: US-08-121-713B-43 (1-8)
 Number of sequences searched: 53402
 Number of scores above cutoff: 3795

Results of the initial comparison of US-08-121-713B-43 (1-8) with:
Data bank : A-Geneseq 18, all entries

100000	N	U50000	M	B	E	R	O	F10000	S	E	5000	Q	U	E	N	C	E	S	1000	500	100	50	10
--------	---	--------	---	---	---	---	---	--------	---	---	------	---	---	---	---	---	---	---	------	-----	-----	----	----

52043

Score	Frequency
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	6
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	0	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	1	2	1.48
Times:	CPU		Total Elapsed
	00:00:31.08		00:00:32.00
Number of residues:		6354270	
Number of sequences searched:		53402	
Number of scores above cutoff:		3795	
Cut-off raised to 3.			
Cut-off raised to 4.			
Cut-off raised to 5.			

The scores below are sorted by initial score.
Significance is calculated based on initial score.
A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Score	Init. Opt.
1. R60334	Blood coagulation inhibiting	5	5	5	2.70
2. R15722	Anticoagulant (5).	5	5	5	2.70
3. R15730	Anticoagulant (13).	6	5	5	2.70
4. R25436	N-terminal auxiliary sequence	8	5	5	2.70
5. R25440	N-terminal auxiliary sequence	14	5	5	2.70
6. R11153	Hirulog-13.	18	5	5	2.70
7. R56242	Fibrinectin-derived peptide.	19	5	7	2.70

8. R26193 D-Cha/D-Npa Hirulog-8. 20 5 5 2.70 0
9. R26192 Hirulog-8. 20 5 5 2.70 0
10. R26395 Sequence of platelet binding 20 5 5 2.70 0
11. R11163 Hirulog-27. 20 5 5 2.70 0
12. R11164 Hirulog-28. 20 5 5 2.70 0
13. R11158 Hirulog-18b. 20 5 5 2.70 0
14. R11151 Hirulog-8. 20 5 5 2.70 0
15. P71352 Synthetic peptide encoding ra 21 5 5 2.70 0
16. R11156 Hirulog-17. 22 5 5 2.70 0
17. R11154 Hirulog-15. 22 5 5 2.70 0
18. R11161 Hirulog-21. 23 5 5 2.70 0
19. R34698 Cys-7 hirulog derivative - an 24 5 7 2.70 0
20. R34697 Native hirulog - anticoagulant 24 5 5 2.70 0
21. R11155 Hirulog-15. 24 5 5 2.70 0
22. R34699 Cys-20 hirulog derivative - a 25 5 5 2.70 0
23. R21422 Matrix peptide from bovine te 31 5 5 2.70 0
24. R05850 Factor VII peptide analogue. 39 5 5 2.70 0
25. R28993 Encoded by human Gs-alpha com 83 5 5 2.70 0
26. R31358 Metastasis inhibiting peptide 90 5 7 2.70 0
27. R25382 Bifunctional inhibitor of pia 96 5 5 2.70 0
28. R28707 Bifunctional inhibitor of pia 96 5 5 2.70 0
29. R25151 Bifunctional inhibitor of pia 96 5 5 2.70 0
30. R28703 Bifunctional inhibitor of pia 96 5 5 2.70 0
31. R25147 Bifunctional inhibitor of pia 97 5 5 2.70 0
32. R28708 Bifunctional inhibitor of pia 97 5 5 2.70 0
33. R28706 Bifunctional inhibitor of pia 97 5 5 2.70 0
34. R28705 Bifunctional inhibitor of pia 97 5 5 2.70 0
35. R28704 Bifunctional inhibitor of pia 97 5 5 2.70 0
36. R13342 NANBH-derived clone GORS22 pr 97 5 5 2.70 0
37. R25154 Bifunctional inhibitor of pia 98 5 5 2.70 0
38. R25150 Bifunctional inhibitor of pia 98 5 5 2.70 0
39. R25153 Bifunctional inhibitor of pia 99 5 5 2.70 0
40. R25149 Bifunctional inhibitor of pia 99 5 5 2.70 0
41. R25152 Bifunctional inhibitor of pia 106 5 5 2.70 0
42. R25148 Bifunctional inhibitor of pia 106 5 5 2.70 0
43. R07997 Human CD2 cytoplasmic domain. 117 5 5 2.70 0
44. R47503 Protein derived from G-CSF st 165 5 5 2.70 0
45. P70247 AA sequence (VII) of a polype 191 5 5 2.70 0

1. US-08-121-713B-43 (1-8)
R60334 Blood coagulation inhibiting peptide.

ID R60334 standard; peptide; 5 AA.
AC R60334;
DT 07-MAR-1995 (first entry)
DE Blood coagulation inhibiting peptide.
KW Blood; fibrin; coagulation; inhibition; thrombin; fibrinogen.
OS Synthetic.
PN J06179696-A.
PD 28-JUN-1994.
PF 19-MAR-1993; 085678.
PR 13-OCT-1992; JP-300380.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
PA (NIHA-) NIPPON HAM KK.
DR WPI; 94-245692/30.
PT New blood coagulation inhibiting peptide(s) having
PT fibrin-agglutination inhibitory activity - useful for the
PT treatment and prevention of thrombosis
PS Claim 1; Page 2; 6pp; Japanese.

CC The blood coagulation inhibiting peptide comprises L-form amino
CC acids and has fibrin agglutination inhibiting activity. Such
CC peptides (see R60327-R60335) were synthesised using a peptide
CC synthesiser (Applied Biosystems Co., 430A) and purified by high
CC pressure liquid chromatography. To 0.2 ml of fibrinogen solution
CC (2.5 mg/ml), 0.2 ml of test peptide solution was added and incubated
CC for 1 minute at 37 deg. Celsius. 0.2 ml of thrombin solution was
CC added and the IC50 was determined according to the method of
CC Kawasaki et al. This peptide had an IC50 of 65 micromolar compared
CC with a control peptide of Gly-Pro-Arg which registered an IC50 of
CC 250 micromolar.
SQ Sequence 5 AA;
SQ 0 A; 1 R; 0 N; 0 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 2 G; 0 H;
SQ 0 I; 0 L; 0 K; 0 M; 0 F; 2 P; 0 S; 0 T; 0 W; 0 Y; 0 V;
Initial Score = 5 Optimized Score = 5 Significance = 2.70
Residue Identity = 80% Matches = 4 Mismatches = 1
Gaps = 0 Conservative Substitutions = 0

X X
PXFRPGXC
||||
GPRPG
X X

2. US-08-121-713B-43 (1-8)
R15722 Anticoagulant (5).

ID R15722 standard; Protein; 5 AA.
AC R15722;
DT 24-JAN-1992 (first entry)
DE Anticoagulant (5).
KW Fibrin; anticoagulant; inhibitor; therapeutic; diagnosis.
OS Synthetic.
FH Key Location/Qualifiers
FT /label= G-NH2, G-NH(Ethyl), G-NH(Ethyl)2,
FT G-NH(Butyl)
PN EP-456152-A.
PD 13-NOV-1991.
PF 06-MAY-1991; 107307.
PR 08-MAY-1990; DE-014655.
PA (BEHW) BEHRINGWERKE AG.
PI Stuber W, Fickenscher K;
DR WPI; 91-334142/46.
PT New anticoagulant penta- and hexa:peptide(s) - are
PT Glycine-Proline-Arginine-Proline amino acid derivs. which inhibit
PT agglomeration of fibrin chains but not thrombin
PS Claim 5; Page 8; 8pp; German.
CC The peptides represented in R15718-33 are anticoagulants which are
CC more effective than previously known chemically similar cpds. They
CC can be used for therapeutic and diagnostic purposes.
SQ Sequence 5 AA;
SQ 0 A; 1 R; 0 N; 0 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 2 G; 0 H;
SQ 0 I; 0 L; 0 K; 0 M; 0 F; 2 P; 0 S; 0 T; 0 W; 0 Y; 0 V;
Initial Score = 5 Optimized Score = 5 Significance = 2.70
Residue Identity = 80% Matches = 4 Mismatches = 1
Gaps = 0 Conservative Substitutions = 0

X X
FXPRPGXC
|||||
GPRPG
X X

3. US-08-121-713B-43 (1-8)
R15730 Anticoagulant (13).

ID R15730 standard; Protein; 6 AA.
AC R15730;
DT 24-JAN-1992 (first entry)
DE Anticoagulant (13).
KW Fibrin; anticoagulant; inhibitor; therapeutic; diagnosis.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc difference 6
FT /label= P-NH2, P-NH(Isopropyl)
FN EP-456152-A.
PD 13-NOV-1991.
PF 06-MAY-1991; 107307.
PR 08-MAY-1990; DE-014655.
PA (BEHW) BEHRINGER AG.
PI Stuber W, Fickenscher K;
DR WPI; 91-334142/46.
PT New anticoagulant penta- and hexa-peptide(s) - are
PT Glycine-Proline-Arginine-Proline amino acid derivs. which inhibit
PT agglomeration of fibrin chains but not thrombin
PS Claim 5; Page 8; 8pp; German.
CC The peptides represented in R15718-33 are anticoagulants which are
CC more effective than previously known chemically similar cpds. They
CC can be used for therapeutic and diagnostic purposes.
SQ Sequence 6 AA;
SQ 0 A; 1 R; 0 N; 0 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 3 G; 0 H;
SQ 0 I; 0 L; 0 K; 0 M; 0 F; 2 P; 0 S; 0 T; 0 W; 0 Y; 0 V;

Initial Score = 5 Optimized Score = 5 Significance = 2.70
Residue Identity = 66% Matches = 4 Mismatches = 2
Gaps = 0 Conservative Substitutions = 0

X X
FXPRPGXC
|||||
GPRPG
X X

4. US-08-121-713B-43 (1-8)
R26436 N-terminal auxiliary sequence.

ID R26436 standard; peptide; 8 AA.
AC R26436;
DT 08-FEB-1993 (first entry)
DE N-terminal auxiliary sequence.
KW Negative hydrophobicity; increased renaturation yield.
OS Synthetic.
FN EP-500108-A.
PD 26-AUG-1992.

PF 20-FEB-1992; 102864.
PR 21-FEB-1991; DE-105480.
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
PI Ambrosius D, Dony C, Rudolph R.
DR WPI; 92-286227/35.
PT Increasing renaturation yield of recombinant protein from
PT prokaryotic host - by expressing it with terminal hydrophilic
PT aminoacid sequence attached, pref. detachable at specific
PT cleavage site
PS Claim 11; Page 14; 18pp; German.
CC The peptide is used as an auxiliary sequence which can be added to
CC the N and/or C-terminus of recombinant proteins which exist in at
CC least partially inactive form and are activated by solubilisation
CC and/or renaturation techniques. The auxiliary sequence has a ratio
CC of relative hydrophobicity: number of amino acids of -2.0 kcal/mole
CC or smaller. Incorporation of the auxiliary sequence increases the
CC yield during the renaturation process. This method is applied to
CC recombinant proteins produced in prokaryotic organisms, especially
CC E. coli. A specific application is production of recombinant
CC granulocyte-colony stimulating factor (GM-CSF) or its derivatives.
CC See also R26437-R26444.

SQ Sequence 8 AA;
SQ 0 A; 1 R; 0 N; 0 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 0 G; 0 H;
SQ 0 I; 1 L; 0 K; 1 M; 0 F; 4 P; 0 S; 1 T; 0 W; 0 Y; 0 V;

Initial Score = 5 Optimized Score = 5 Significance = 2.70
Residue Identity = 66% Matches = 4 Mismatches = 2
Gaps = 0 Conservative Substitutions = 0

X X
FXPRPGXC
|||||
MTPLPRPP
X X

5. US-08-121-713B-43 (1-8)
R26440 N-terminal auxiliary sequence.

ID R26440 standard; peptide; 14 AA.
AC R26440;
DT 08-FEB-1993 (first entry)
DE N-terminal auxiliary sequence.
KW Negative hydrophobicity; increased renaturation yield.
OS Synthetic.
FN EP-500108-A.
PD 26-AUG-1992.

PF 20-FEB-1992; 102864.
PR 21-FEB-1991; DE-105480.
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
PI Ambrosius D, Dony C, Rudolph R.
DR WPI; 92-286227/35.

PT Increasing renaturation yield of recombinant protein from
PT prokaryotic host - by expressing it with terminal hydrophilic
PT aminoacid sequence attached, pref. detachable at specific
PT cleavage site
PS Claim 11; Page 14; 18pp; German.
CC The peptide is used as an auxiliary sequence which can be added to
CC the N and/or C-terminus of recombinant proteins which exist in at
CC least partially inactive form and are activated by solubilisation

CC and/or renaturation techniques. The auxiliary sequence has a ratio
 CC of relative hydrophobicity: number of amino acids of -2.0 kcal/mole
 CC or smaller. Incorporation of the auxiliary sequence increases the
 CC yield during the renaturation process. This method is applied to
 CC recombinant proteins produced in prokaryotic organisms, especially
 CC E. coli. A specific application is production of recombinant
 CC granulocyte-colony stimulating factor (GM-CSF) or its derivatives.
 CC See also R26436-R26444.
 SQ Sequence 14 AA;
 SQ 0 A; 1 R; 0 N; 0 D; 0 B; 0 C; 0 Q; 2 E; 0 Z; 1 G; 0 H;
 SQ 0 I; 2 L; 0 K; 1 M; 0 F; 5 P; 0 S; 2 T; 0 W; 0 Y; 0 V;

Initial Score = 5 Optimized Score = 2.70
 Residue Identity = 66% Matches = 4 Mismatches = 2
 Gaps = 0 Conservative Substitutions = 0

X X
 XPRFGXC
 | | |
 MTPLEGTPLPRPP
 10 X

6. US-08-121-713B-43 (1-8)
 R1153 Hirulog-13.

ID R1153 standard; Protein; 18 AA.

AC R1153;

DT 21-MAY-1991 (first entry)

DE Hirulog-13.

KW Hirudin; thrombin inhibitor.

OS Synthetic.

FH Key Location/Qualifiers

FT Modified -site 1

FT /label= D-Phe

PN W09102750-A.

PD 07-MAR-1991.

PF 17-AUG-1990; U04642.

PR 18-AUG-1989; US-395482.

PR 06-JUL-1990; US-549388.

PA (BIOG-) BIOGEN INC.

PA (HEAL-) HEALTH RES. INC.

PI Maraganore JM, Fenton JW, Kline T;

DR WPI; 91-087245/12.

PT Thrombin inhibitor for therapeutic, prophylactic purposes - used

PT to treat or prevent vascular disease, inflammatory responses,

PT carcinoma(s), neuro-degenerative and chronic thromboembolic

PT diseases

PS Example 16; Page 50; 125pp; English.

CC The peptide mimics the action of hirudin and inhibits thrombin. It

CC is smaller and more potent than hirudin, so less likely to provoke

CC an immune response. The peptide can be used to prevent/treat

CC vascular diseases, inflammation, carcinomas and neurodegenerative

CC diseases. It can also be used for ex vivo thrombus imaging, for

CC storing and treating extracorporeal blood and for coating invasive

CC devices. R1151-R1156.

CC See also R18 AA;

SQ Sequence 18 AA;

SQ 0 A; 1 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 4 E; 0 Z; 3 G; 0 H;

SQ 1 I; 1 L; 0 K; 0 M; 2 F; 3 P; 0 S; 0 T; 0 W; 1 Y; 0 V;

Initial Score = 5 Optimized Score = 2.70
 Residue Identity = 57% Matches = 4 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 XPRFGXC
 | | | |
 FPRFGNGDFEEIPEEY
 X X 10

7. US-08-121-713B-43 (1-8)
 R56242 Fibrinectin-derived peptide.

ID R56242 standard; peptide; 19 AA.

AC R56242;

DT 12-JAN-1995 (first entry)

DE Fibrinectin-derived peptide.

KW Fibrinectin; inflammation; antiinflammatory; immunosuppressive;

KW leukocyte; arthritis; autoimmune disease; graft versus host disease.

OS Synthetic.

PN W09413692-A.

PD 23-JUN-1994.

PF 06-DEC-1993; U11781.

PR 10-DEC-1992; US-990296.

PR 21-OCT-1993; US-139903.

PA (MIND) UNIV MINNESOTA.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Allen JB, Furcht LT, McCarthy JB, Wahl SM;

DR WPI; 94-217799/26.

PT Fibrinectin derived peptide(s) for treating inflammation -

PT involving leucocyte activation, partic. arthritis and graft

PT versus host disease

PS Claim 6; Page 30; 45pp; English.

CC Chronic inflammation or autoimmune diseases are treated with peptides

CC corresponding to residues 1906-24 (R56242), 1946-60 (R56243), 1892-99

CC (R56244), 1961-85 (R56245), 1784-92 (R56246) and 1485-504 (R56247) of

CC fibrinectin.

SQ Sequence 19 AA;

SQ 0 A; 3 R; 0 N; 0 D; 0 B; 0 C; 0 Q; 2 E; 0 Z; 0 G; 0 H;

SQ 0 I; 0 L; 1 K; 0 M; 0 F; 6 P; 1 S; 0 T; 0 W; 1 Y; 3 V;

Initial Score = 5 Optimized Score = 2.70
 Residue Identity = 71% Matches = 5 Mismatches = 2
 Gaps = 0 Conservative Substitutions = 0

X X
 XPRFGXC
 | | | |
 KPGSPPREWVRPRPGV
 10 X X

8. US-08-121-713B-43 (1-8)
 R26193 D-Cha/D-Npa Hirulog-8.

ID R26193 standard; peptide; 20 AA.

AC R26193;

DT 09-FEB-1993 (first entry)

DE D-Cha/D-Npa Hirulog-8.
 KW Thrombin; inhibitor; hirudin; metastasis; adult respiratory;
 KW distress; neurodegenerative; Alzheimer's; Parkinson's; intravascular;
 KW coagulation; imaging; reperfusion; tPA.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc difference 1
 FT /note= "D-form"
 FT naphthyl alanine"
 PN WO9213952-A.
 PD 20-AUG-1992.
 PF 03-FEB-1992; D00836.
 PR 08-FEB-1991; US-652929.
 PA (BIOJ) BIOGEN INC.
 PI Bourdon PR, Maraganore JM, Jablonski JM;
 DR WPI; 92-300038/36.
 PT New catalytic site-directed thrombin inhibitors - used for
 PT treatment and prevention of thrombotic disorders,
 PT thrombin-induced inflammation, carcinoma(s), etc.
 PS Example 4; Page 36; 57pp; English.
 CC D-Cha/D-Npa Hirulog-8 are examples of catalytic site directed
 CC thrombin inhibitors. The inhibitors are designed for optimum
 CC spatial configuration so are more active than hirudin (which
 CC they mimic qualitatively) and can be used at higher doses. Also
 CC they are less likely to be immunogenic. Particular applications
 CC include inhibition of metastatic tumours, treatment and prevention
 CC of thrombin-induced inflammation (e.g. in adult respiratory distress
 CC syndrome or reperfusion injury), to treat neurodegenerative diseases
 CC (e.g. Alzheimer's or Parkinson's diseases), to inhibit thrombus
 CC accretion and platelet dependent thrombosis and to treat or prevent
 CC disseminated intravascular coagulation. The inhibitors can also be
 CC used, in conjunction with a thrombolytic agent, e.g. tPA, to decrease
 CC reperfusion time and increase reocclusion time. The inhibitors can
 CC also be used in extracorporeal blood circulations, to coat the surface
 CC of invasive devices, and (when coupled to a radioisotope such as iodine
 CC 123) for ex vivo imaging of a fibrin or platelet thrombus.
 CC See also R26192-4.
 SQ Sequence 20 AA;
 SQ 0 A; 1 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 4 E; 0 Z; 5 G; 0 H;
 SQ 1 I; 1 L; 0 K; 0 M; 1 F; 3 P; 0 S; 0 T; 0 W; 1 Y; 0 V;
 SQ 1 Others;

Initial Score = 5 Optimized Score = 5 Significance = 2.70
 Residue Identity = 71% Matches = 5 Mismatches = 2
 Gaps = 0 Conservative Substitutions = 0

X X
 FXPRPGXC
 |||||
 XPRPGGGNGDFEIPF
 X X 10

9. US-08-121-713B-43 (1-8)
 R26192 Hirulog-8.

ID R26192 standard; peptide; 20 AA.
 AC R26192;
 DT 09-FEB-1993 (first entry)
 DE Hirulog-8.

KW Thrombin; inhibitor; hirudin; metastasis; adult respiratory;
 KW distress; neurodegenerative; Alzheimer's; Parkinson's; intravascular;
 KW coagulation; imaging; reperfusion; tPA.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc difference 1
 FT /note= "D-form"
 FT WO9213952-A.
 PD 20-AUG-1992.
 PF 03-FEB-1992; U00836.
 PR 08-FEB-1991; US-652929.
 PA (BIOJ) BIOGEN INC.
 PI Bourdon PR, Maraganore JM, Jablonski JM;
 DR WPI; 92-300038/36.
 PT New catalytic site-directed thrombin inhibitors - used for
 PT treatment and prevention of thrombotic disorders,
 PT thrombin-induced inflammation, carcinoma(s), etc.
 PS Example 2; Page 33; 57pp; English.
 CC Hirulog-8 is an example of a catalytic site directed thrombin
 CC inhibitor. The inhibitors are designed for optimum spatial
 CC configuration so are more active than hirudin (which they mimic
 CC qualitatively) and can be used at higher doses. Also they are less
 CC likely to be immunogenic. Particular applications include of
 CC inhibition of metastatic tumours, treatment and prevention of
 CC thrombin-induced inflammation (e.g. in adult respiratory distress
 CC syndrome or reperfusion injury), to treat neurodegenerative diseases
 CC (e.g. Alzheimer's or Parkinson's diseases), to inhibit thrombus
 CC accretion and platelet dependent thrombosis and to treat or prevent
 CC disseminated intravascular coagulation. The inhibitors can also be
 CC used, in conjunction with a thrombolytic agent, e.g. tPA, to decrease
 CC reperfusion time and increase reocclusion time. The inhibitors can
 CC also be used in extracorporeal blood circulations, to coat the surface
 CC of invasive devices, and (when coupled to a radioisotope such as iodine
 CC 123) for ex vivo imaging of a fibrin or platelet thrombus.
 CC See also R26193-4.
 SQ Sequence 20 AA;
 SQ 0 A; 1 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 4 E; 0 Z; 5 G; 0 H;
 SQ 1 I; 1 L; 0 K; 0 M; 2 F; 3 P; 0 S; 0 T; 0 W; 1 Y; 0 V;
 SQ 1 Others;

Initial Score = 5 Optimized Score = 5 Significance = 2.70
 Residue Identity = 57% Matches = 4 Mismatches = 3
 Gaps = 0 Conservative Substitutions = 0

X X
 FXPRPGXC
 |||||
 FXPRPGGGNGDFEIPF
 X X 10

10. US-08-121-713B-43 (1-8)
 R26395 Sequence of platelet binding peptide.

ID R26395 standard; peptide; 20 AA.
 AC R26395;
 DT 25-JAN-1993 (first entry)
 DE Sequence of platelet binding peptide.
 KW Technetium-99m labelled polypeptide imaging agent;
 OS radiolabeled imaging; radiodiagnostic agent.
 OS Synthetic.

FH Key Location/Qualifiers
FT Modified site 1
FT /label= D-Phe
PN W09213572-A.
PD 20-AUG-1992.
PF 07-FEB-1992; U00757.
PR 08-FEB-1991; US-653012.
PA (DIAT-) DIATECH INC.
PI Dean RT;
DR WPI; 92-299767/36.
PT New technetium-99m labelled polypeptide imaging agents - for
PT imaging of clots, tumours, infection sites, atherosclerotic and
PT amyloid plaques or bone, and for visualising organs
PS Claim 6; Page 13; 19pp; English.
CC The binding peptide is covalently linked to a 'Cp(aa)Cp' technetium
CC binding group wherein Cp is a protected cysteine and (aa) is an amino
CC acid. The technetium-99m complexes are used to image target sites
CC within a mammalian body.
SQ Sequence 20 AA;
SQ 0 A; 1 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 4 E; 0 Z; 5 G; 0 H;
SQ 1 I; 1 L; 0 K; 0 M; 2 F; 3 P; 0 S; 0 T; 0 W; 1 Y; 0 V;
Initial Score = 5 Optimized Score = 5 Significance = 2.70
Residue Identity = 57% Matches = 4 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XPXPGXC
||||
FPRPGGGNGDFEEIPE
X X 10

11. US-08-121-713B-43 (1-8)
R1163 Hirulog-27.

ID R1163 standard; Protein; 20 AA.
AC R1163;
DT 21-MAY-1991 (first entry)
DE Hirulog-27.
KW Hirudin; thrombin inhibitor.
OS Synthetic.
FH Key Location/Qualifiers
FT Modified -site 1
FT /label= D-Phe
FT Modified -site 3
FT /label= OTHER
FT /note= "Arg-(CO-CH2)"
PN W09102750-A.
PD 07-MAR-1991.
PF 17-AUG-1990; U04642.
PR 18-AUG-1989; US-395482.
PR 06-JUL-1990; US-549388.
PA (BIOG-) BIOGEN INC.
PA (HEAL-) HEALTH RES INC.
PI Maraganore JM, Fenton JW, Kline T;
DR WPI; 91-087245/12.
PT Thrombin inhibitor for therapeutic, prophylactic purposes - used
PT to treat or prevent vascular disease, inflammatory responses,
PT carcinoma(s), neuro-degenerative and chronic thromboembolic

PT diseases
PS Example 26; Page 67; 125pp; English.
CC The peptide mimics the action of hirudin and inhibits thrombin. It
CC is smaller and more potent than hirudin, so less likely to provoke
CC an immune response. The peptide can be used to prevent/treat
CC vascular diseases, inflammation, carcinomas and neurodegenerative
CC diseases. It can also be used for ex vivo thrombus imaging, for
CC storing and treating extracorporeal blood and for coating invasive
CC devices.
CC See also R11151-R11166.
SQ Sequence 20 AA;
SQ 0 A; 1 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 4 E; 0 Z; 5 G; 0 H;
SQ 1 I; 1 L; 0 K; 0 M; 2 F; 3 P; 0 S; 0 T; 0 W; 1 Y; 0 V;
Initial Score = 5 Optimized Score = 5 Significance = 2.70
Residue Identity = 57% Matches = 4 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XPXPGXC
||||
FPRPGGGNGDFEEIPE
X X 10

12. US-08-121-713B-43 (1-8)
R1164 Hirulog-28.

ID R1164 standard; Protein; 20 AA.
AC R1164;
DT 21-MAY-1991 (first entry)
DE Hirulog-28.
KW Hirudin; thrombin inhibitor.
OS Synthetic.
FH Key Location/Qualifiers
FT Modified -site 1
FT /label= D-Phe
FT Modified -site 3
FT /label= OTHER
FT /note= "Arg-(CH2N)"
PN W09102750-A.
PD 07-MAR-1991.
PF 17-AUG-1990; U04642.
PR 18-AUG-1989; US-395482.
PR 06-JUL-1990; US-549388.
PA (BIOG-) BIOGEN INC.
PA (HEAL-) HEALTH RES INC.
PI Maraganore JM, Fenton JW, Kline T;
DR WPI; 91-087245/12.
PT Thrombin inhibitor for therapeutic, prophylactic purposes - used
PT to treat or prevent vascular disease, inflammatory responses,
PT carcinoma(s), neuro-degenerative and chronic thromboembolic
PT diseases
PS Example 27; Page 68; 125pp; English.
CC The peptide mimics the action of hirudin and inhibits thrombin. It
CC is smaller and more potent than hirudin, so less likely to provoke
CC an immune response. The peptide can be used to prevent/treat
CC vascular diseases, inflammation, carcinomas and neurodegenerative
CC diseases. It can also be used for ex vivo thrombus imaging, for
CC storing and treating extracorporeal blood and for coating invasive

CC devices.
CC See also R11151-R11166.
SQ Sequence 20 AA;
SQ 0 A; 1 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 4 E; 0 Z; 5 G; 0 H;
SQ 1 I; 1 L; 0 K; 0 M; 2 F; 3 P; 0 S; 0 T; 0 W; 1 Y; 0 V;
Initial Score = 5 Optimized Score = 5 Significance = 2.70
Residue Identity = 57% Matches = 4 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XPRPGXC
||||
FPRGGGNGDFEEIPE
X X 10

13. US-08-121-713B-43 (1-8)
R11158 Hirulog-18b.

ID R11158 standard; Protein; 20 AA.
AC R11158;
DT 21-MAY-1991 (first entry)
DE Hirulog-18b.
KW Hirudin; thrombin inhibitor.
OS Synthetic.
FH Key Location/Qualifiers
FT Modified -site 1
FT /label= D-Phe
FT Modified -site 3
FT /label= OTHER
FT /note= "beta-homoarginine"
EN W09102750-A.
PD 07-MAR-1991.
PF 17-AUG-1990; 004642.
PR 18-AUG-1989; US-395482.
PR 06-JUL-1990; US-549388.
PA (BIOG-) BIOGEN INC.
PA (HEAL-) HEALTH RES INC.
PI Maraganore JM, Fenton JW, Kline T;
DR WPI; 91-087245/12.
PT Thrombin inhibitor for therapeutic, prophylactic purposes - used
PT to treat or prevent vascular disease, inflammatory responses,
PT carcinoma(s), neuro-degenerative and chronic thromboembolic
PT diseases
PS Example 21; Page 52; 125pp; English.
CC The peptide mimics the action of hirudin and inhibits thrombin. It
CC is smaller and more potent than hirudin, so less likely to provoke
CC an immune response. The peptide can be used to prevent/treat
CC vascular diseases, inflammation, carcinomas and neurodegenerative
CC diseases. It can also be used for ex vivo thrombus imaging, for
CC storing and treating extracorporeal blood and for coating invasive
CC devices.
CC See also R11151-R11166.
SQ Sequence 20 AA;
SQ 0 A; 1 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 4 E; 0 Z; 5 G; 0 H;
SQ 1 I; 1 L; 0 K; 0 M; 2 F; 3 P; 0 S; 0 T; 0 W; 1 Y; 0 V;

Initial Score = 5 Optimized Score = 5 Significance = 2.70
Residue Identity = 57% Matches = 4 Mismatches = 3

Gaps = 0 Conservative Substitutions = 0

X X
XPRPGXC
||||
FPRGGGNGDFEEIPE
X X 10

14. US-08-121-713B-43 (1-8)
R11151 Hirulog-8.

ID R11151 standard; peptide; 20 AA.
AC R11151;
DT 21-MAY-1991 (first entry)
DE Hirulog-8.
KW Hirudin; thrombin inhibitor; Hirulog-9; Hirulog-11; Hirulog-12.
OS Synthetic.
FH Key Location/Qualifiers
FT Modified -site 1
FT /label= D-Phe
FT W09102750-A.
PD 07-MAR-1991.
PF 17-AUG-1990; 004642.
PR 18-AUG-1989; US-395482.
PR 06-JUL-1990; US-549388.
PA (BIOG-) BIOGEN INC.
PA (HEAL-) HEALTH RES INC.
PI Maraganore JM, Fenton JW, Kline T;
DR WPI; 91-087245/12.
PT Thrombin inhibitor for therapeutic, prophylactic purposes - used
PT to treat or prevent vascular disease, inflammatory responses,
PT carcinoma(s), neuro-degenerative and chronic thromboembolic
PT diseases

Example 4; Page 39; 125pp; English.
CC The peptide mimics the action of hirudin and inhibits thrombin. It
CC is smaller and more potent than hirudin, so less likely to provoke
CC an immune response. The peptide can be used to prevent/treat
CC vascular diseases, inflammation, carcinomas and neurodegenerative
CC diseases. It can also be used for ex vivo thrombus imaging, for
CC storing and treating extracorporeal blood and for coating invasive
CC devices. The same sequence but with a D-Pro at posn. 4 replacing
CC the L-Pro is designated Hirulog-9; Hirulog-8 with the Tyr at posn.
CC 19 modified to 3,5-diiodoTyr is known as Hirulog-11, and modified
CC to Tyr(OSO3) as Hirulog-12.
CC See also R11152-R11166.

SQ Sequence 20 AA;
SQ 0 A; 1 R; 1 N; 1 D; 0 B; 0 C; 0 Q; 4 E; 0 Z; 5 G; 0 H;
SQ 1 I; 1 L; 0 K; 0 M; 2 F; 3 P; 0 S; 0 T; 0 W; 1 Y; 0 V;

Initial Score = 5 Optimized Score = 5 Significance = 2.70
Residue Identity = 57% Matches = 4 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XPRPGXC
||||
FPRGGGNGDFEEIPE
X X 10


```

15. US-08-121-713B-43 (1-8)
   P71352 Synthetic peptide encoding rabies virus coat glyco
ID P71352 standard; protein; 21 AA.
AC P71352;
DT 01-JAN-1980 (first entry)
DE Synthetic peptide encoding rabies virus coat glycoprotein antigenic
DE determinant.
KW Rabies virus; glycoprotein; coat protein; vaccine; antigen;
KW immunogen; mammal; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc_difference 12
FT /label= N or T
FT Misc_difference 20
FT /label= N or K
FT W08700179-A.
EN 15-JAN-1987.
PD 03-JUL-1986; U01424.
PR 03-JUL-1985; US-752222.
PR 12-DEC-1986; US-941163.
PA (SALK ) SALK INST FOR BIOL STUD.
PI Patrick JW, Heinemann SF, Boss BD, Cowan WM;
DR WPI; 87-021980/03.
PT Synthetic peptide-based anti-rabies compns. - having an amino
PT acid sequence of a segment of the rabies virus coat glyco:protein
PS Claim 2; Page 53; 58pp; English.
CC When this peptide is conjugated to a carrier protein which is
CC immunogenic in a mammal, it is capable of raising an immune
CC response against rabies virus, and may therefore be used as a
CC potential anti-rabies vaccine. The peptide may also be used in the
CC treatment, prophylaxis and diagnosis of rabies. See also P71350-1
CC P71353-55.
SQ Sequence 21 AA;
SQ 1 A; 3 R; 0 N; 1 D; 0 B; 1 C; 0 Q; 0 E; 0 Z; 3 G; 0 H;
SQ 1 I; 0 L; 1 K; 0 M; 1 F; 3 P; 2 S; 2 T; 0 W; 0 Y; 0 V;
SQ 2 Others;

Initial Score = 5 Optimized Score = 5 Significance = 2.70
Residue Identity = 66% Matches = 4 Mismatches = 2
Gaps = 0 Conservative Substitutions = 0

X X
XPXPRGXG
|||||
X X 10
PRGTPCDIFTXSRGK
> 0 <
> 0 <
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sq43pir.res made by on Fri 19 May 95 8:49:14-PDT.

Query sequence being compared:US-08-121-713B-43 (1-8)
Number of sequences searched: 75511

```

PARAMETERS

Similarity matrix Unitary 1 K-tuple 2
Mismatch penalty 1.00 Joining penalty 20
Gap penalty 0.05 Window size 6
Gap size penalty 0
Cutoff score 0
Randomization group 0
Initial scores to save 45 Alignments to save 15
Optimized scores to save 0 Display context 10

SEARCH STATISTICS

Scores: Mean 2 Median 4 Standard Deviation 1.29

Times: CPU 00:01:12.06 Total Elapsed 00:01:13.00

Number of residues: 22468834
Number of sequences searched: 75511
Number of scores above cutoff: 4067

Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

The scores below are sorted by initial score.
Significance is calculated based on initial score.
A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Init. Opt.	Sig.
1. A27797	class I histocompatibility antigen HLA alpha chain (clone RS5) precursor - human	356	7	7	3.86
2. JQ0532	probable substance P - small	753	7	7	3.86
3. S33300	bradykinin-potentiating peptidase (EC 3.1.1.1)	11	5	5	2.32
4. A37196	angiotensin-converting enzyme (EC 2.7.4.8)	13	5	5	2.32
5. XAV19B	T cell receptor J-alpha w1.2	24	5	5	2.32
6. S40139	acetylcholinesterase (EC 3.1.1.7)	47	5	5	2.32
7. S27042	bactenecin 7 - bovine	59	5	5	2.32
8. A36589	acetylcholinesterase (EC 3.1.1.7)	72	5	5	2.32
9. B41573	oncoprotein zL-Myc - zebra fi	77	5	5	2.32
10. E48059	progesterone receptor, form B	91	5	5	2.32
11. A61552	guanylate kinase (EC 2.7.4.8)	98	5	5	2.32
12. S39448	hypothetical hydroxylproline-r	100	5	5	2.32
13. B38277	VI protein - tobacco yellow d	102	5	5	2.32
14. A42452	hypothetical protein 103 - ph	103	5	5	2.32
15. S20822	hypothetical protein 103 - ph	103	5	5	2.32
16. S15151	hypothetical protein - Rhizob	113	5	5	2.32
17. S02365	GTP cyclohydrolase I (EC 3.5.5.1)	114	5	5	2.32
18. S48846					

19. PN0139	lutropin beta chain - minke w	118	5	5	2.32
20. PN0141	lutropin beta chain - sperm w	118	5	5	2.32
21. S34165	keratin type II, hair - sheep	123	5	5	2.32
22. S20178	hypothetical protein (IFM1 3'	125	5	5	2.32
23. JU0068	homeotic protein Chox1 - chic	126	5	5	2.32
24. B48662	hypothetical protein repA4 -	128	5	5	2.32
25. Q0ECB1	hypothetical protein RepA4 -	128	5	5	2.32
26. Q0ECBR	hypothetical protein RepA4 -	128	5	5	2.32
27. S12371	hsp protein - human	130	5	5	2.32
28. A26839	ORF1 protein - Escherichia co	134	5	5	2.32
29. S01387	U1 snRNP protein C - human	159	5	5	2.32
30. A44263	U1-specific snRNP C protein -	159	5	5	2.32
31. S09404	Mannose 6-phosphate receptor	163	5	5	2.32
32. S14549	DNA-directed RNA polymerase (165	5	5	2.32
33. JC2081	granulocyte colony-stimulat	165	5	5	2.32
34. A42361	DNA-directed RNA polymerase (165	5	5	2.32
35. JH0656	hypothetical 18.3K protein (n	168	5	5	2.32
36. S36749	transcription factor HES-3 -	175	5	5	2.32
37. S24298	chorion protein - silkworm	178	5	5	2.32
38. S37398	adenine phosphoribosyltransfe	182	5	5	2.32
39. JK0219	somatotropin - African elepha	190	5	5	2.32
40. KIPGGU	guanylate kinase (EC 2.7.4.8)	197	5	5	2.32
41. S39447	guanylate kinase (EC 2.7.4.8)	198	5	5	2.32
42. S14456	XA-1 protein - African clawed	198	5	5	2.32
43. S32545	guanylate kinase - pig	199	5	5	2.32
44. S43041	guanylate kinase (EC 2.7.4.8)	207	5	5	2.32
45. KIECGU	guanylate kinase (EC 2.7.4.8)	207	5	5	2.32

1. US-08-121-713B-43 (1-8)

class I histocompatibility antigen HLA alpha chain

ENTRY A27797 #type complete
TITLE class I histocompatibility antigen HLA alpha chain (clone RS5) precursor - human
ORGANISM #formal name Homo sapiens #common name man
DATE 18-Oct-1989 #sequence_revision 30-Jun-1991 #text_change 18-Jun-1993
ACCESSIONS A27797
REFERENCE A27797
#authors Srivastava, R.; Chorney, M.J.; Lawrence, S.K.; Pan, J.; Smith, Z.; Smith, C.L.; Weissman, S.M.
#journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:4224-4228
#title Structure, expression, and molecular mapping of a divergent member of the class I HLA gene family.
#cross-references M01D:87231978
#accession A27797
#molecule type DNA
##residues 1-356 ##label SRI
##cross-references GB:M16714

GENETICS 21/3; 110/3; 202/1; 333/1; 343/3
#introns #length 356 #molecular-weight 39484 #checksum 7384
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.86
Residue Identity = 62% Matches = 5 Mismatches = 3
Gaps 0 Conservative Substitutions
X X

XPXPRPGXC
| | | | |
TLLLLSEARPLPPRGSGSHLKYIPLS
10 X 20 X 30

2. US-08-121-713B-43 (1-8)
JQ0532 OP protein - Kennedy yellow mosaic virus

ENTRY JQ0532 #type complete
TITLE OP protein - Kennedy yellow mosaic virus
ORGANISM #formal_name Kennedy yellow mosaic virus
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
JQ0532 31-Dec-1993

ACCESSIONS JQ0532
REFERENCE JQ0532
#authors Ding, S.; Keese, P.; Gibbs, A.
#journal J. Gen. Virol. (1990) 71:925-931
#title The nucleotide sequence of the genomic RNA of Kennedy yellow mosaic tymovirus-Jervis Bay isolate: relationships with potex- and carlaviruses.

#contents Strain Jervis Bay isolate
#accession JQ0532
#molecule_type mRNA
#residues 1-753 #label DIN

SUMMARY #length 753 #molecular_weight 82427 #checksum 9812
SEQUENCE

Initial Score = 7 Optimized Score = 7 Significance = 3.86
Residue Identity = 62% Matches = 5 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XPXPRPGXC
| | | | |
AASSQSSSILPLPPRGNGRPGVLPKVR
260 270 280

3. US-08-121-713B-43 (1-8)
S33300 probable substance P - smaller spotted catshar

ENTRY S33300 #type complete
TITLE probable substance P - smaller spotted catshar
ORGANISM #formal_name Scyliorhinus canicula #common_name smaller spotted catshark, smaller spotted dogfish
DATE 08-Dec-1993; #sequence_revision 08-Dec-1993; #text_change
S33300 08-Dec-1993

ACCESSIONS S33300
REFERENCE S33300
#authors Waugh, D.; Wang, Y.; Hazon, N.; Balment, R.J.; Conlon, J.M.
#journal Eur. J. Biochem. (1993) 214:469-474
#title Primary structures and biological activities of substance-P-related peptides from the brain of the dogfish, Scyliorhinus canicula.

#accession S33300
#status preliminary
#residues 1-11 #label WAU

SUMMARY #length 11 #molecular_weight 1278 #checksum 4938
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 57% Matches = 4 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XPXPRPGXC
| | | | |
KPRPGQFFGLM
X X 10

4. US-08-121-713B-43 (1-8)
A37196 bradykinin-potentiating peptide 1 - island jararac

ENTRY A37196 #type complete
TITLE bradykinin-potentiating peptide 1 - island jararaca
ORGANISM #formal_name Bothrops insularis #common_name island jararaca
DATE 14-Feb-1992 #sequence_revision 01-Dec-1992 #text_change
05-Aug-1994

ACCESSIONS A37196
REFERENCE A37196
#authors Cintra, A.C.O.; Vieira, C.A.; Giglio, J.R.
#journal J. Protein Chem. (1990) 9:221-227
#title Primary structure and biological activity of bradykinin potentiating peptides from Bothrops insularis snake venom.

#cross-references MUID:90351557
#accession A37196
#status preliminary
#molecule_type protein
#residues 1-13 #label CIN

KEYWORDS pyroglutamic acid
FEATURE 1

#modified_site pyrrolidone carboxylic acid (Gln) #status experimental

SUMMARY #length 13 #molecular_weight 1388 #checksum 7017
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 50% Matches = 4 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XPXPRPGXC
| | | | |
QGGWPRGPEIPP
X X 10

5. US-08-121-713B-43 (1-8)
XAV19B angiotensin-converting enzyme inhibitor V-9 - jara

ENTRY XAV19B #type complete
TITLE angiotensin-converting enzyme inhibitor V-9 - jararaca
ORGANISM #formal_name Bothrops jararaca #common_name jararaca
DATE #sequence_revision 13-Jul-1981 #text_change 08-Dec-1994
A01253
ACCESSIONS A90356
REFERENCE Ondetti, M.A.; Williams, N.J.; Sabo, E.F.; Pluscec, J.; Weaver, E.R.; Kocy, O.

#journal Biochemistry (1971) 10:4033-4039
#title Angiotensin-converting enzyme inhibitors from the venom of
Bothrops jararaca. Isolation, elucidation of structure, and
synthesis.
#cross-references MUID:72118526
#accession A01253
##molecule_type protein
##residues 1-13 ##label OND
##note the structure of the peptide was confirmed by synthesis
COMMENT This peptide also potentiates bradykinin by inhibiting the kinases
that inactivate it.
CLASSIFICATION #superfamily bradykinin-potentiating peptide
KEYWORDS pyroglutamic acid
FEATURE
1 #modified site pyrrolidone carboxylic acid (Gln) #status
experimental
SUMMARY #length 13 #molecular-weight 1388 #checksum 7017
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 50% Matches = 4 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XPRPGXC
|||||
QGGWPRPGPEIPP
X 10

6. US-08-121-713B-43 (1-8)

S40139 T cell receptor J-alpha wNI.2 - human

ENTRY S40139 #type complete
TITLE T cell receptor J-alpha wNI.2 - human
ORGANISM #formal name Homo sapiens #common name man
DATE 07-Dec-1994; #sequence_revision 07-Dec-1994; #text_change
07-Dec-1994
ACCESSIONS S40139
REFERENCE S40133
#authors Plaza, A.; Kono, D.H.; Theofilopoulos, A.N.
#submission submitted to the EMBL Data Library, February 1993
#accession S40139
##status preliminary
##residues 1-24 ##label PLA
##cross-references EMBL:X71025
SUMMARY #length 24 #molecular-weight 2380 #checksum 3180
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 50% Matches = 4 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XPRPGXC
|||||
VSPRPGAGNKLTFGGGT
X 10

7. US-08-121-713B-43 (1-8)

S27042 acetylcholinesterase (EC 3.1.1.7) - rat
ENTRY S27042 #type complete
TITLE acetylcholinesterase (EC 3.1.1.7) - rat
ORGANISM #formal name Rattus norvegicus #common name Norway rat
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993
ACCESSIONS S27042
REFERENCE S27042
#authors Legay, C.; Bon, S.; Massoulié, J.
#journal FEBS Lett. (1993) 315:163-166
#title Expression of a cDNA encoding the glycolipid-anchored form of
rat acetylcholinesterase.

#accession S27042
##status preliminary
##residues 1-47 ##label IEG
SUMMARY #length 47 #molecular-weight 4960 #checksum 5424
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 50% Matches = 4 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
XPRPGXC
|||||
CTCPSPAHGEEAPRGPALSLSLFFLF
X 30

8. US-08-121-713B-43 (1-8)

A36589 bactenecin 7 - bovine

ENTRY A36589 #type complete
TITLE bactenecin 7 - bovine
ORGANISM #formal name Bos primigenius taurus #common name cattle
DATE 12-Apr-1991 #sequence_revision 12-Apr-1991 #text_change
24-Jun-1993
ACCESSIONS A36589
REFERENCE A36589
#authors Frank, R.W.; Gennaro, R.; Schneider, K.; Przybylski, M.;
Romeo, D.
#journal J. Biol. Chem. (1990) 265:18871-18874
#title Amino acid sequences of two proline-rich bacterenecins.
Antimicrobial peptides of bovine neutrophils.

#cross-references MUID:91035404
#accession A36589

##status preliminary
##molecule_type protein
##residues_1-59 ##label FRA
SUMMARY #length 59 #molecular-weight 6910 #checksum 9841
SEQUENCE

Initial Score = 5 Optimized Score = 7 Significance = 2.32
Residue Identity = 62% Matches = 5 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
XPRPGXC

|||||
RLPRPRPLPFPGRPRIPRLPFP
10 20 X 30

9. US-08-121-713B-43 (1-8)
B41573 acetylcholinesterase (EC 3.1.1.7) - mouse (fragment)
ENTRY #type fragment
TITLE acetylcholinesterase (EC 3.1.1.7) - mouse (fragment)
ORGANISM #formal name Mus musculus #common name house mouse
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 23-Jun-1993
ACCESSIONS B41573
REFERENCE A41573
#authors Li, Y.; Camp, S.; Rachinsky, T.L.; Getman, D.; Taylor, P.
#journal J. Biol. Chem. (1991) 266:23083-23090
#title Gene structure of mammalian acetylcholinesterase. Alternative exons dictate tissue-specific expression.
#cross-references MUID:92078174
#accession B41573
#status preliminary
#molecule_type DNA
#residues 1-72 #label LIA
#cross-references GB:M76540
KEYWORDS carboxylic ester hydrolase
SUMMARY #length 72 #checksum 2652
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 50% Matches = 4 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
X
PXPRPGXC

|||||
CTCFSPAHEAARPGDIALSLILFLF
10 X 20 X 30

10. US-08-121-713B-43 (1-8)
E48059 oncoprotein zL-Myc - zebra fish (fragment)

ENTRY #type fragment
TITLE oncoprotein zL-Myc - zebra fish (fragment)
ORGANISM #formal name Brachydanio rerio #common name zebra fish
DATE 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
ACCESSIONS E48059
REFERENCE A48059
#authors Schreiber-Agus, N.; Horner, J.; Torres, R.; Chiu, F.C.; DePinho, R.A.

#journal Mol. Cell. Biol. (1993) 13:2765-2775
#title Zebra fish myc family and max genes: differential expression and oncogenic activity throughout vertebrate evolution.
#cross-references MUID:93233639
#accession E48059
#status preliminary
#molecule_type mRNA; DNA
#residues 1-77 #label SCH

#cross-references NCBI:129812
#note sequence extracted from NCBI backbone
#note sequence not compared to nucleotide translation
SUMMARY #length 77 #checksum 9459
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 50% Matches = 4 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
X
PXPRPGXC

|||||
SPSRTLGDWLPFRPGDRWGAAEGLTC
20 X 30 X 40

11. US-08-121-713B-43 (1-8)
A61552 progesterone receptor, form B - chicken (fragments)

ENTRY #type fragments
TITLE progesterone receptor, form B - chicken (fragments)
ORGANISM #formal name Gallus gallus #common name chicken
DATE 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 15-Oct-1994
ACCESSIONS A61552
REFERENCE A61552
#authors Simpson, R.J.; Grego, B.; Govindan, M.V.; Gronemeyer, H.
#journal Mol. Cell. Endocrinol. (1987) 52:177-184
#title Peptide sequencing of the chick oviduct progesterone receptor form B.
#accession A61552
#status preliminary
#molecule_type protein
#residues 1-91 #label SIM

SUMMARY #length 91 #checksum 8063
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 50% Matches = 4 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
X
PXPRPGXC

|||||
VDAGPAGPQSQPRGAAAPGPDADAGP
10 X 20

12. US-08-121-713B-43 (1-8)
S39448 guanylate kinase (EC 2.7.4.8) - bovine

ENTRY #type complete
TITLE guanylate kinase (EC 2.7.4.8) - bovine
ORGANISM #formal name Bos primigenius taurus #common name cattle
DATE 20-May-1994 #sequence_revision 20-May-1994 #text_change 20-May-1994
ACCESSIONS S39448
REFERENCE S39448
#authors Gaidarov, I.O.; Suslov, O.N.; Abdulaev, N.G.

#journal FEBS Lett. (1993) 335:81-84
#title Enzymes of the cyclic GMP metabolism in bovine retina. I.
#accession S39448
#status preliminary
#residues 1-98 #label GAI
#length 98 #molecular-weight 10747 #checksum 8606

SUMMARY
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 50% Matches = 4 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

X X
PXPRPGXC
|||||
GSIFSVSHTTRDPRPGENGKDYFVTR
20 X 30 40

13. US-08-121-713B-43 (1-8)
B38277 hypothetical hydroxyproline-rich glycoprotein (clo

ENTRY B38277 #type fragment
TITLE hypothetical hydroxyproline-rich glycoprotein (clone MG1.6a)
- Chlamydomonas reinhardtii (fragment)
ORGANISM #formal_name Chlamydomonas reinhardtii
DATE 31-May-1991 #sequence_revision 31-May-1991 #text_change
30-Sep-1993

ACCESSIONS B38277
REFERENCE A38277
#authors Adair, W.S.; Apt, K.E.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:7355-7359
#title Cell wall regeneration in Chlamydomonas: accumulation of
mRNAs encoding cell wall hydroxyproline-rich glycoproteins.
#cross-references M01D:91017504

#accession B38277
#status preliminary
#molecule type mRNA
#residues 1-100 #label ADA
#note sequence not compared to nucleotide translation
SUMMARY #length 100 #checksum 4244
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 50% Matches = 4 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
PXPRPGXC
|||||
RPLQPERLVRAPRRPGAPRQOQLPLP
80 X 90 100

14. US-08-121-713B-43 (1-8)
A42452 V1 protein - tobacco yellow dwarf virus (strain Au)
ENTRY A42452 #type complete
TITLE V1 protein - tobacco yellow dwarf virus (strain Australia)

ORGANISM #formal_name tobacco yellow dwarf virus
DATE 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change
ACCESSIONS A42452
REFERENCE 08-Apr-1994
#authors Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.;
Thomas, J.E.
#journal Virology (1992) 187:633-642
#title The nucleotide sequence of the infectious cloned DNA
component of tobacco yellow dwarf virus reveals features of
geminiviruses infecting monocotyledonous plants.

#accession A42452
#molecule type DNA
#residues 1-102 #label MOR
#cross-references GB:M81103
SUMMARY #length 102 #molecular-weight 11178 #checksum 6879
SEQUENCE

Initial Score = 5 Optimized Score = 5 Significance = 2.32
Residue Identity = 50% Matches = 4 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

X X
PXPRPGXC
|||||
TTTEICGNTPLRRPGEGNPGGPV
80 90 X 100

15. US-08-121-713B-43 (1-8)

S20822 hypothetical protein 103 - phage Pfi

ENTRY S20822 #type complete
TITLE hypothetical protein 103 - phage Pfi
ORGANISM #formal_name phage Pfi
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993

ACCESSIONS S20822
REFERENCE S20696
#authors Hill, D.F.; Short, N.J.; Perham, R.N.; Petersen, G.B.
#submission submitted to the EMBL Data Library, March 1990
#accession S20822
#status preliminary
#residues 1-103 #label HIL

SUMMARY #cross-references EMBL:X52107
#length 103 #molecular-weight 11154 #checksum 8403
SEQUENCE

Initial Score = 5 Optimized Score = 7 Significance = 2.32
Residue Identity = 62% Matches = 5 Mismatches = 3
Gaps = 0 Conservative Substitutions = 0

X X
PXPRPGXC
|||||
NHGRQLLTPRPRGLRLLCVLRKAG
10 X 20